

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 30, 2004, 13:22:14 ; Search time 34 Seconds  
(without alignments)  
238.386 Million cell updates/sec

Title: US-10-025-598-2

Perfect score: 190

Sequence: 1 MDWLKARVEQLOALFARGTDSNAELRAMEAKLKAEIQK 39

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 789580 seqs, 207824079 residues

Total number of hits satisfying chosen parameters: 789580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications AA:

- 1: /cgn2\_6/prodata/2/pubpaa/US07\_PUBCOMB.pep.\*
- 2: /cgn2\_6/prodata/2/pubpaa/PCT\_NEW\_PUB.pep.\*
- 3: /cgn2\_6/prodata/2/pubpaa/US06\_NEW\_PUB.pep.\*
- 4: /cgn2\_6/prodata/2/pubpaa/US06\_PUBCOMB.pep.\*
- 5: /cgn2\_6/prodata/2/pubpaa/US07\_NEW\_PUB.pep.\*
- 6: /cgn2\_6/prodata/2/pubpaa/PCTUS\_PUBCOMB.pep.\*
- 7: /cgn2\_6/prodata/2/pubpaa/US08\_NEW\_PUB.pep.\*
- 8: /cgn2\_6/prodata/2/pubpaa/US08\_PUBCOMB.pep.\*
- 9: /cgn2\_6/prodata/2/pubpaa/US09A\_PUBCOMB.pep.\*
- 10: /cgn2\_6/prodata/2/pubpaa/US09B\_PUBCOMB.pep.\*
- 11: /cgn2\_6/prodata/2/pubpaa/US09C\_PUBCOMB.pep.\*
- 12: /cgn2\_6/prodata/2/pubpaa/US09\_NEW\_PUB.pep.\*
- 13: /cgn2\_6/prodata/2/pubpaa/US10A\_PUBCOMB.pep.\*
- 14: /cgn2\_6/prodata/2/pubpaa/US10B\_PUBCOMB.pep.\*
- 15: /cgn2\_6/prodata/2/pubpaa/US10C\_PUBCOMB.pep.\*
- 16: /cgn2\_6/prodata/2/pubpaa/US10\_NEW\_PUB.pep.\*
- 17: /cgn2\_6/prodata/2/pubpaa/US60\_NEW\_PUB.pep.\*
- 18: /cgn2\_6/prodata/2/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	190	100.0	39	12	US-10-025-598-2
2	190	100.0	39	12	US-10-326-441-2
3	85	44.7	17	12	US-10-210-023-111
4	60.5	31.8	1549	12	US-10-369-493-5702
5	59.5	31.3	49	15	US-10-023-066A-30
6	59.5	31.3	49	15	US-10-023-066A-32
7	59.5	31.3	49	15	US-10-023-066A-54
8	59.5	31.3	56	15	US-10-023-066A-79
9	59.5	31.3	77	15	US-10-023-066A-75
10	58.5	30.8	263	10	US-09-925-300-1468
11	58	30.5	30	12	US-09-911-261A-19
12	58	30.5	30	12	US-10-173-551-39
13	58	30.5	30	15	US-10-057-408-19
14	58	30.5	45	9	US-09-766-378A-33
15	57.5	30.3	42	15	US-10-023-066A-34

16	57.5	30.3	44	11	US-09-491-614-31	Sequence 31, Appl
17	57.5	30.3	56	15	US-10-023-066A-77	Sequence 77, Appl
18	57	30.0	341	12	US-10-192-381-32	Sequence 32, Appl
19	56	29.5	361	9	US-09-841-132-299	Sequence 299, App
20	56	29.5	647	9	US-09-841-132-436	Sequence 436, App
21	55.5	29.2	73	9	US-09-864-761-33395	Sequence 33395, A
22	55.5	29.2	2099	15	US-10-128-714-3290	Sequence 3290, Ap
23	55.5	29.2	2405	15	US-10-128-714-8290	Sequence 8290, Ap
24	55	28.9	28	15	US-10-023-066A-70	Sequence 70, Appl
25	55	28.9	31	10	US-09-756-983-19	Sequence 19, Appl
26	55	28.9	194	12	US-10-211-689-30	Sequence 30, Appl
27	55	28.9	364	10	US-09-756-983-22	Sequence 22, Appl
28	55	28.9	445	15	US-10-158-761-14752	Sequence 14752, A
29	55	28.9	545	12	US-09-849-602-25	Sequence 25, Appl
30	55	28.9	572	12	US-09-738-630-105	Sequence 105, App
31	54.5	28.7	222	9	US-09-840-787-41	Sequence 41, Appl
32	54.5	28.7	663	12	US-10-369-493-19140	Sequence 19140, A
33	54.5	28.7	854	11	US-09-770-107-2	Sequence 2, Appli
34	54	28.4	74	9	US-09-864-761-45412	Sequence 45412, A
35	54	28.4	454	12	US-10-369-493-23078	Sequence 23078, A
36	54	28.4	660	12	US-09-841-260-139	Sequence 139, App
37	54	28.4	660	14	US-10-007-693-139	Sequence 139, App
38	54	28.4	1091	12	US-10-369-493-6328	Sequence 6328, App
39	54	28.4	2383	15	US-10-082-830-260	Sequence 260, App
40	54	28.4	5795	9	US-09-815-242-12610	Sequence 12610, A
41	53.5	28.2	1118	12	US-10-369-493-18845	Sequence 18845, A
42	53.5	28.2	2288	12	US-10-369-493-6774	Sequence 6774, Ap
43	53.5	28.2	3500	15	US-10-153-219-2	Sequence 2, Appli
44	53.5	28.2	3537	15	US-10-153-219-15	Sequence 15, Appl
45	53	27.9	151	9	US-09-840-787-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1

US-10-025-598-2  
; Sequence 2, Application US/10025598  
; Publication No. US20030147852A1  
; GENERAL INFORMATION:  
; APPLICANT: Schack, Diane L.  
; TITLE OF INVENTION: TOXIN-PHAGE BACTERIOCID ANTIBIOTIC AND USES THEREOF  
; FILE REFERENCE: 13688-002001  
; CURRENT APPLICATION NUMBER: US/10/025,598  
; CURRENT FILING DATE: 2002-04-08  
; NUMBER OF SEQ ID NOS: 16  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 2  
; LENGTH: 39  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Synthetically generated peptide  
US-10-025-598-2

Query Match 100.0%; Score 190; DB 12; Length 39;

Best Local Similarity 100.0%; Pred. No. 6.1e-18;

Matches 39; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MDWLKARVEQLOALFARGTDSNAELRAMEAKLKAEIQK 39  
|||||  
Db 1 MDWLKARVEQLOALFARGTDSNAELRAMEAKLKAEIQK 39

RESULT 2

US-10-326-441-2  
; Sequence 2, Application US/10326441  
; Publication No. US20030175246A1  
; GENERAL INFORMATION:  
; APPLICANT: Schack, Diane L.  
; TITLE OF INVENTION: TOXIN-PHAGE BACTERIOCID ANTIBIOTIC AND USES THEREOF  
; FILE REFERENCE: 00246-269001

Search Notes  
10/025,598  
2/21/04

```
; CURRENT APPLICATION NUMBER: US/10/326,441
; CURRENT FILING DATE: 2002-12-18
; PRIOR APPLICATION NUMBER: US 10/025,598
; PRIOR FILING DATE: 2001-12-18
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: Fast-Seq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 39
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetically generated peptide
US-10-326-441-2

Query Match      100.0%; Score 190; DB 12; Length 39;
Best Local Similarity 100.0%; Pred. No. 6.1e-18;
Matches 39; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MDWLKARVEQLQALEARGTDSNAELRAMEAKLKAEIQK 39
   |||||
Db 1 MDWLKARVEQLQALEARGTDSNAELRAMEAKLKAEIQK 39

RESULT 3
US-10-210-023-111
; Sequence 111, Application US/10210023
; Publication No. US20030190740A1
; GENERAL INFORMATION:
; APPLICANT: Altman, Elliot
; TITLE OF INVENTION: STABILIZED BIOACTIVE PEPTIDES AND METHODS OF IDENTIFICATION, SYNTHESIS AND USE
; FILE REFERENCE: 235.00010120
; CURRENT APPLICATION NUMBER: US/10/210,023
; CURRENT FILING DATE: 2002-11-12
; PRIOR APPLICATION NUMBER: 09/701,947
; PRIOR FILING DATE: 2000-12-05
; PRIOR APPLICATION NUMBER: PCT/US99/23731
; PRIOR FILING DATE: 1999-10-12
; PRIOR APPLICATION NUMBER: 60/104,013
; PRIOR FILING DATE: 1998-10-13
; PRIOR APPLICATION NUMBER: 60/112,150
; PRIOR FILING DATE: 1998-12-14
; NUMBER OF SEQ ID NOS: 115
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 111
; LENGTH: 17
; TYPE: PRT
; ORGANISM: ARTIFICIAL
; FEATURE:
; OTHER INFORMATION: a-helical moiety
US-10-210-023-111

Query Match      44.7%; Score 85; DB 12; Length 17;
Best Local Similarity 100.0%; Pred. No. 0.00014;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 DWLKVQEQLQALEAR 18
   |||||
Db 1 DWLKVQEQLQALEAR 17

RESULT 4
US-10-369-493-5702
; Sequence 5702, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; OTHER INFORMATION: PLANTS WITH IMPROVED PROPERTIES
```

```
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 5702
; LENGTH: 1549
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
US-10-369-493-5702

Query Match      31.8%; Score 60.5; DB 12; Length 1549;
Best Local Similarity 36.1%; Pred. No. 40;
Matches 13; Conservative 10; Mismatches 10; Indels 3; Gaps 1;

QY 7 RVQEQLQALEARG---TDSNAELRAMEAKLKAEIQK 39
   |||||
Db 990 RFEKQIKKLTQKGEENTKQSELSAAETKLEGLKK 1025

RESULT 5
US-10-023-066A-30
; Sequence 30, Application US/10023066A
; Publication No. US20030056242A1
; GENERAL INFORMATION:
; APPLICANT: E. I. DU PONT DE NEMOURS AND COMPANY
; TITLE OF INVENTION: CHIMERIC GENES AND METHODS FOR INCREASING THE LYSINE AND THREONINE CONTENT OF THE SEEDS OF PLANTS
; NUMBER OF SEQUENCES: 107
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: E. I. DU PONT DE NEMOURS AND COMPANY
; STREET: 1007 MARKET STREET
; CITY: WILMINGTON
; STATE: DELAWARE
; COUNTRY: U.S.A.
; ZIP: 19898
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: MICROSOFT WORD VERSION 2.0C
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/023,066A
; FILING DATE: 29-Apr-2002
; CLASSIFICATION: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: BARBARA C. SIEGELL
; REGISTRATION NUMBER: 30,684
; REFERENCE/DOCKET NUMBER: BB-1037-C
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 302-992-4931
; TELEFAX: 302-773-0164
; TELEX: 835420
; INFORMATION FOR SEQ ID NO: 30:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 49 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 30:
US-10-023-066A-30

Query Match      31.3%; Score 59.5; DB 15; Length 49;
Best Local Similarity 44.4%; Pred. No. 1;
Matches 16; Conservative 9; Mismatches 10; Indels 1; Gaps 1;

QY 4 LKARVEQLQALEARGTDSNAELRAMEAKLKAEIQK 39
   |||||
```

```
, STREET: 1007 MARKET STREET
, CITY: WILMINGTON
, STATE: DELAWARE
, COUNTRY: U.S.A.
, ZIP: 19898
, COMPUTER READABLE FORM:
, MEDIUM TYPE: FLOPPY DISK
, COMPUTER: IBM PC COMPATIBLE
, OPERATING SYSTEM: PC-DOS/MS-DOS
, SOFTWARE: MICROSOFT WORD VERSION 2.0C
, CURRENT APPLICATION DATA:
, APPLICATION NUMBER: US/10/023,066A
, FILING DATE: 29-Apr-2002
, CLASSIFICATION: <Unknown>
, ATTORNEY/AGENT INFORMATION:
, NAME: BARBARA C. SIEGELL
, REGISTRATION NUMBER: 30,684
, REFERENCE/DOCKET NUMBER: BB-1037-C
, TELECOMMUNICATION INFORMATION:
, TELEPHONE: 302-992-4931
, TELEFAX: 302-773-0164
, TELEX: 835420
, INFORMATION FOR SEQ ID NO: 54:
, SEQUENCE CHARACTERISTICS:
, LENGTH: 49 amino acids
, TYPE: amino acid
, TOPOLOGY: linear
, MOLECULE TYPE: protein
, SEQUENCE DESCRIPTION: SEQ ID NO: 54:
US-10-023-066A-54

Query Match      31.3%; Score 59.5; DB 15; Length 49;
Best Local Similarity 44.4%; Pred. No. 1;
Matches 16; Conservative 9; Mismatches 10; Indels 1

Qy   4 LKA-RVQEQLQALEARGTDSNAELRAMEAKLKAEIQK 39
    ||| :||:::||: :||::||| ::||| :|
Db   5 LKA-MEEKLKAEMEKLKAMEEKLKAMEEKLKAMEEK 39

RESULT 8
US-10-023-066A-79
; Sequence 79, Application US/10023066A
; Publication No. US20030056242A1
; GENERAL INFORMATION:
; APPLICANT: E. I. DU PONT DE NEMOURS AND COMPANY
; TITLE OF INVENTION: CHIMERIC GENES AND METHODS FOR INCREASING THE LYXINE AND THREONINE CONTENT OF THE SEEDS OF PLANTS
; NUMBER OF SEQUENCES: 107
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: E. I. DU PONT DE NEMOURS AND COMPANY
; STREET: 1007 MARKET STREET
; CITY: WILMINGTON
; STATE: DELAWARE
; COUNTRY: U.S.A.
; ZIP: 19898
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: MICROSOFT WORD VERSION 2.0C
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/023,066A
; FILING DATE: 29-Apr-2002
; CLASSIFICATION: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: BARBARA C. SIEGELL
; REGISTRATION NUMBER: 30,684
; REFERENCE/DOCKET NUMBER: BB-1037-C
```

## TELECOMMUNICATION INFORMATION:

TELEPHONE: 302-992-4931  
TELEFAX: 302-773-0164  
TELEX: 835420

## INFORMATION FOR SEQ ID NO: 79:

SEQUENCE CHARACTERISTICS:  
LENGTH: 56 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
SEQUENCE DESCRIPTION: SEQ ID NO: 79:  
US-10-023-066A-79

Query Match 31.3%; Score 59.5; DB 15; Length 56;  
Best Local Similarity 44.4%; Pred. No. 1.2;  
Matches 16; Conservative 9; Mismatches 10; Indels 1; Gaps 1;

QY 4 LKARVEQLQALFARGTDSNAELRAMEAKLKABIQK 39

Db 5 LKA-MEEKLKAMEEKLKAMEEKLKAMEEK 39

## RESULT 9

US-10-023-066A-75

Sequence 75, Application US/10023066A  
Publication No. US20030056242A1

## GENERAL INFORMATION:

APPLICANT: E. I. DU PONT DE NEMOURS AND  
COMPANY

TITLE OF INVENTION: CHIMERIC GENES AND METHODS FOR  
INCREASING THE LYSINE AND  
THREONINE CONTENT OF THE SEEDS OF  
PLANTS

NUMBER OF SEQUENCES: 107

CORRESPONDENCE ADDRESS:

ADDRESSEE: E. I. DU PONT DE NEMOURS

AND COMPANY

STREET: 1007 MARKET STREET

CITY: WILMINGTON

STATE: DELAWARE

COUNTRY: U.S.A.

ZIP: 19898

COMPUTER READABLE FORM:

MEDIUM TYPE: FLOPPY DISK

COMPUTER: IBM PC COMPATIBLE

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: MICROSOFT WORD VERSION 2.0C

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/10/023,066A

FILING DATE: 29-Apr-2002

CLASSIFICATION: <unknown>

ATTORNEY/AGENT INFORMATION:

NAME: BARBARA C. SIEGELL

REGISTRATION NUMBER: 30,684

REFERENCE/DOCKET NUMBER: BB-1037-C

TELECOMMUNICATION INFORMATION:

TELEPHONE: 302-992-4931

TELEFAX: 302-773-0164

TELEX: 835420

INFORMATION FOR SEQ ID NO: 75:

SEQUENCE CHARACTERISTICS:

LENGTH: 77 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

SEQUENCE DESCRIPTION: SEQ ID NO: 75:

US-10-023-066A-75

Query Match 31.3%; Score 59.5; DB 15; Length 77;  
Best Local Similarity 44.4%; Pred. No. 1.7;  
Matches 16; Conservative 9; Mismatches 10; Indels 1; Gaps 1;

QY 4 LKARVEQLQALFARGTDSNAELRAMEAKLKABIQK 39

Db 5 LKA-MEEKLKAMEEKLKAMEEKLKAMEEK 39

## RESULT 10

US-09-925-300-1468

Sequence 1468, Application US/09925300

Patent No. US20020151681A1

GENERAL INFORMATION:

APPLICANT: Craig Rosen,

APPLICANT: Steve Ruben

TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies

FILE REFERENCE: PA101

CURRENT APPLICATION NUMBER: US/09/925,300

CURRENT FILING DATE: 2001-08-10

PRIOR APPLICATION NUMBER: PCT/US00/05988

PRIOR FILING DATE: 2000-03-08

PRIOR APPLICATION NUMBER: 60/124,270

PRIOR FILING DATE: 1999-03-12

NUMBER OF SEQ ID NOS: 1890

SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 1468

LENGTH: 263

TYPE: PRT

ORGANISM: Homo sapiens

FEATURE:

NAME/KEY: SITE

LOCATION: (6)

OTHER INFORMATION: xaa equals any of the naturally occurring L-amino acids  
US-09-925-300-1468

Query Match 30.8%; Score 58.5; DB 10; Length 263;

Best Local Similarity 36.8%; Pred. No. 9.4;

Matches 14; Conservative 11; Mismatches 12; Indels 1; Gaps 1;

QY 2 DMLKARVEQLQALFARGT-DSNAELRAMEAKLKABIQ 38

Db 80 EFLEKKIEQLTAAKKHGTNKRAALQALKRKRVEKQ 117

## RESULT 11

US-09-911-261A-19

Sequence 19, Application US/09911261A

Publication No. US20030134350A1

GENERAL INFORMATION:

APPLICANT: Sera Takashi

TITLE OF INVENTION: Zinc Finger Domain Recognition Code and Uses Thereof

FILE REFERENCE: 109845.135

CURRENT APPLICATION NUMBER: US/09/911,261A

CURRENT FILING DATE: 2001-07-23

PRIOR APPLICATION NUMBER: US 60/220,060

PRIOR FILING DATE: 2000-07-21

NUMBER OF SEQ ID NOS: 69

SOFTWARE: PatentIn version 3.0

SEQ ID NO 19

LENGTH: 30

TYPE: PRT

ORGANISM: Artificial Sequence

FEATURE:

OTHER INFORMATION: Acid dimerization peptide.

US-09-911-261A-19

Query Match 30.5%; Score 58; DB 12; Length 30;

Best Local Similarity 46.4%; Pred. No. 0.9;

Matches 13; Conservative 6; Mismatches 9; Indels 0; Gaps 0;

QY 6 ARVEQLQALFARGTDSNAELRAMEAKL 33

Db 1 AQLEKELQALEKENAQLEWELQALEKEL 28

## RESULT 12

US-10-173-551-39

```
; Sequence 39, Application US/10173551
; Publication No. US20030232387A1
; GENERAL INFORMATION:
; APPLICANT: Lu, Chafen
; TITLE OF INVENTION: Antibodies that bind alphaE Integrin
; FILE REFERENCE: 1855 2025-000
; CURRENT APPLICATION NUMBER: US/10/173,551
; CURRENT FILING DATE: 2002-06-14
; NUMBER OF SEQ ID NOS: 70
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 39
; LENGTH: 30
; TYPE: PRT
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: acidic peptide
US-10-173-551-39

Query Match          30.5%; Score 58; DB 12; Length 30;
Best Local Similarity 46.4%; Pred. No. 0.9;
Matches 13; Conservative 6; Mismatches 9; Indels 0; Gaps 0;

QY      6 ARVEQLQALEARGTDSNAELRAMEAKL 33
DB      1 AQLEKELQALEKENAQLEWELQALEKEL 28
      :::::|||||
      :::::|||||

RESULT 13
US-10-057-408-19
; Sequence 19, Application US/10057408
; Publication No. US20030082561A1
; GENERAL INFORMATION:
; APPLICANT: Sera, Takashi
; TITLE OF INVENTION: Zinc Finger Domain Recognition Code and Uses Thereof
; FILE REFERENCE: 109845 135
; CURRENT APPLICATION NUMBER: US/10/057,408
; CURRENT FILING DATE: 2002-01-23
; PRIOR APPLICATION NUMBER: US 60/220,060
; PRIOR FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 69
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 19
; LENGTH: 30
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Acid dimerization peptide.
US-10-057-408-19

Query Match          30.5%; Score 58; DB 15; Length 30;
Best Local Similarity 46.4%; Pred. No. 0.9;
Matches 13; Conservative 6; Mismatches 9; Indels 0; Gaps 0;

QY      6 ARVEQLQALEARGTDSNAELRAMEAKL 33
DB      1 AQLEKELQALEKENAQLEWELQALEKEL 28
      :::::|||||
      :::::|||||

RESULT 14
US-09-766-378A-33
; Sequence 33, Application US/09766378A
; Patent No. US20020091079A1
; GENERAL INFORMATION:
; APPLICANT: Rhode, Peter R.
; Acevedo, Jorge
; Burkhardt, Martin
; Jiao, Jin-an
; Wong, Hing C.
; TITLE OF INVENTION: SOLUBLE MHC COMPLEXES AND METHODS OF USE THEREOF
; NUMBER OF SEQUENCES: 38
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dike, Bronstein, Roberts & Cushman, LLP
```

```
; STREET: 130 Water Street
; CITY: Boston
; STATE: MA
; COUNTRY: usa
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/766,378A
; FILING DATE: 19-Jan-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/960,190
; FILING DATE: 29-OCT-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Corless, Peter F.
; REGISTRATION NUMBER: 33,860
; REFERENCE/DOCKET NUMBER: 48002-DIV
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-523-3400
; TELEFAX: 617-523-6440
; TELEX: <Unknown>
; INFORMATION FOR SEQ ID NO: 33:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 45 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 33:
US-09-766-378A-33

Query Match          30.5%; Score 58; DB 9; Length 45;
Best Local Similarity 46.4%; Pred. No. 1.4;
Matches 13; Conservative 6; Mismatches 9; Indels 0; Gaps 0;

QY      6 ARVEQLQALEARGTDSNAELRAMEAKL 33
DB      16 AQLEKELQALEKENAQLEWELQALEKEL 43
      :::::|||||
      :::::|||||

RESULT 15
US-10-023-066A-34
; Sequence 34, Application US/10023066A
; Publication No. US20030056242A1
; GENERAL INFORMATION:
; APPLICANT: E. I. DU PONT DE NEMOURS AND COMPANY
; TITLE OF INVENTION: CHIMERIC GENES AND METHODS FOR INCREASING THE LYSINE AND THREONINE CONTENT OF THE SEEDS OF PLANTS
; NUMBER OF SEQUENCES: 107
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: E. I. DU PONT DE NEMOURS AND COMPANY
; STREET: 1007 MARKET STREET
; CITY: WILMINGTON
; STATE: DELAWARE
; COUNTRY: U.S.A.
; ZIP: 19898
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: MICROSOFT WORD VERSION 2.0C
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/023,066A
; FILING DATE: 29-Apr-2002
; CLASSIFICATION: <Unknown>
```

```

; ATTORNEY/AGENT INFORMATION:
; NAME: BARBARA C. SIEGELL
; REGISTRATION NUMBER: 30,684
; REFERENCE/DOCKET NUMBER: BB-1037-C
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 302-992-4931
; TELEFAX: 302-773-0164
; TELEX: 835420
; INFORMATION FOR SEQ ID NO: 34:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 42 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 34:
US-10-023-066A-34

Query Match 30.3%; Score 57.5; DB 15; Length 42;
Best Local Similarity 41.7%; Pred. No. 1.5;
Matches 15; Conservative 10; Mismatches 10; Indels 1; Gaps 1;

QY 4 LKARVEQELQALEARGTDSNAELRAMEAKLKAKIQK 39
Db 5 MKA-MEEKLKAMEEKLKAMEEKLKAMEEKLKAMEEK 39

```

Search completed: January 30, 2004, 13:27:45  
Job time : 34 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 30, 2004, 13:20:14 ; Search time 22 Seconds  
(without alignments)  
75.006 Million cell updates/sec

Title: us-10-025-598-2

Perfect score: 190

Sequence: 1 MDWLKARVEQLQALEARGTDSNAELRAMEAKLKAEIQK 39

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

- Database : Issued Patents AA:\*
- 1: /cgn2\_6/prodata/1/iaa/5A COMB.pep.\*
  - 2: /cgn2\_6/prodata/1/iaa/5B COMB.pep.\*
  - 3: /cgn2\_6/prodata/1/iaa/6A COMB.pep.\*
  - 4: /cgn2\_6/prodata/1/iaa/6B COMB.pep.\*
  - 5: /cgn2\_6/prodata/1/iaa/PCTUS COMB.pep.\*
  - 6: /cgn2\_6/prodata/1/iaa/backfile1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	59.5	31.3	49	1 US-08-182-175A-33	Sequence 33, Appl
2	59.5	31.3	49	1 US-08-182-175A-35	Sequence 35, Appl
3	59.5	31.3	49	1 US-08-182-175A-45	Sequence 45, Appl
4	59.5	31.3	49	1 US-08-474-633A-30	Sequence 30, Appl
5	59.5	31.3	49	1 US-08-474-633A-32	Sequence 32, Appl
6	59.5	31.3	49	1 US-08-474-633A-54	Sequence 54, Appl
7	59.5	31.3	49	4 US-08-823-771-30	Sequence 30, Appl
8	59.5	31.3	49	4 US-08-823-771-32	Sequence 32, Appl
9	59.5	31.3	49	4 US-08-823-771-54	Sequence 54, Appl
10	59.5	31.3	49	5 PCT-US92-06412-33	Sequence 33, Appl
11	59.5	31.3	49	5 PCT-US92-06412-35	Sequence 35, Appl
12	59.5	31.3	49	5 PCT-US92-06412-45	Sequence 45, Appl
13	59.5	31.3	56	1 US-08-182-175A-91	Sequence 91, Appl
14	59.5	31.3	56	1 US-08-182-175A-112	Sequence 112, Appl
15	59.5	31.3	56	1 US-08-474-633A-79	Sequence 79, Appl
16	59.5	31.3	56	4 US-08-823-771-79	Sequence 79, Appl
17	59.5	31.3	56	5 PCT-US92-06412-91	Sequence 91, Appl
18	59.5	31.3	56	5 PCT-US92-06412-112	Sequence 112, Appl
19	59.5	31.3	77	1 US-08-182-175A-57	Sequence 57, Appl
20	59.5	31.3	77	1 US-08-474-633A-75	Sequence 75, Appl
21	59.5	31.3	77	4 US-08-823-771-75	Sequence 75, Appl
22	59.5	31.3	77	5 PCT-US92-06412-57	Sequence 57, Appl
23	58	30.5	45	3 US-08-960-190A-33	Sequence 33, Appl
24	57.5	30.3	42	1 US-08-182-175A-37	Sequence 37, Appl
25	57.5	30.3	42	1 US-08-474-633A-34	Sequence 34, Appl
26	57.5	30.3	42	4 US-08-823-771-34	Sequence 34, Appl
27	57.5	30.3	42	5 PCT-US92-06412-37	Sequence 37, Appl

28	57.5	30.3	56	1 US-08-182-175A-59	Sequence 59, Appl
29	57.5	30.3	56	1 US-08-474-633A-77	Sequence 77, Appl
30	57.5	30.3	56	4 US-08-823-771-77	Sequence 77, Appl
31	57.5	30.3	56	5 PCT-US92-06412-59	Sequence 59, Appl
32	57	30.0	220	4 US-09-328-352-5166	Sequence 5166, Ap
33	56	29.5	361	4 US-09-556-877-299	Sequence 299, App
34	56	29.5	361	4 US-09-620-412C-299	Sequence 299, App
35	56	29.5	361	4 US-09-598-419-299	Sequence 299, App
36	55.5	29.2	167	4 US-09-107-532A-6197	Sequence 6197, Ap
37	55.5	29.2	284	3 US-09-053-197A-2	Sequence 2, Appli
38	55.5	29.2	284	4 US-09-085-761A-2	Sequence 7, Appli
39	55.5	29.2	584	1 US-08-179-738-7	Sequence 7, Appli
40	55.5	29.2	584	2 US-08-628-145-7	Sequence 5, Appli
41	55.5	29.2	591	1 US-08-179-738-5	Sequence 10, Appl
42	55.5	29.2	591	1 US-08-179-738-10	Sequence 5, Appli
43	55.5	29.2	591	2 US-08-628-145-5	Sequence 10, Appl
44	55.5	29.2	591	2 US-08-628-145-10	Sequence 16, Appl
45	55.5	29.2	595	1 US-08-171-718-16	

ALIGNMENTS

RESULT 1  
US-08-182-175A-33  
; Sequence 33, Application US/08182175A  
; Patent No. 5559223  
; GENERAL INFORMATION:  
; APPLICANT: Saverio Carl Falco  
; APPLICANT: Sharon J. Keeler  
; TITLE OF INVENTION: Synthetic Storage Proteins with Defined Structure Containing P  
; NUMBER OF SEQUENCES: 113  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: E. I. du Pont de Nemours and Company  
; STREET: 1007 Market Street  
; CITY: Wilmington  
; STATE: Delaware  
; COUNTRY: USA  
; ZIP: 19898  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: Macintosh  
; OPERATING SYSTEM: Macintosh System, 6.0  
; SOFTWARE: Microsoft Word, 4.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/182,175A  
; FILING DATE:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 07/743,006  
; FILING DATE: 9 August 1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Linda Axamechy Floyd  
; REGISTRATION NUMBER: 33,692  
; REFERENCE/DOCKET NUMBER: BB-1031  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (302) 992-4929  
; TELEFAX: (302) 892-7949  
; TELEX: 835420  
; INFORMATION FOR SEQ ID NO: 33:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 49 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-182-175A-33

Query Match 31.3%; Score 59.5; DB 1; Length 49;  
Best Local Similarity 44.4%; Pred. No. 0.41; Indels 1; Gaps 1;  
Matches 16; Conservative 9; Mismatches 10;

QY 4 LKARVEQLQALEARGTDSNAELRAMEAKLKAEIQK 39

10/025,598  
Search Notes  
2/2/04





1. The first step is to identify the problem or question that needs to be answered. This involves understanding the context and the specific requirements of the task.

AND COMPANY



```

; TITLE OF INVENTION: Synthetic Storage Proteins with Defined Structure Containing P
; NUMBER OF SEQUENCES: 113
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: E.I. du Pont de Nemours and Company
; STREET: 1007 Market Street
; CITY: Wilmington
; STATE: Delaware
; COUNTRY: USA
; ZIP: 19898
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy Disk
; COMPUTER: Macintosh
; OPERATING SYSTEM: Macintosh System, 6.0
; SOFTWARE: Microsoft Word, 4.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US92/06412
; FILING DATE: 19920807
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/743,006
; FILING DATE: 9 August 1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Linda Axamethy Floyd
; REGISTRATION NUMBER: 33,692
; REFERENCE/DOCKET NUMBER: BB-1031
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (302) 992-4929
; TELEFAX: (302) 892-7949
; TELEX: 835420
; INFORMATION FOR SEQ ID NO: 35:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 49 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; PCT-US92-06412-35

Query Match      31.3%; Score 59.5; DB 5; Length 49;
Best Local Similarity 44.4%; Pred. No. 0.41;
Matches 16; Conservative 9; Mismatches 10; Indels 1; Gaps 1;

Qy   4 LKARVEDELQALEARGTDSNAELRAMEAKLKAEIQK 39
     ||| :||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db   12 LKA-MEEKLAMESKLMAMEBKLMAMEBKLMAMEEK 46

RESULT 12
PCT-US92-06412-45
; Sequence 45, Application PC/TUS9206412
; GENERAL INFORMATION:
; APPLICANT: Saverio Carl Falco
; APPLICANT: Sharon J. Keeler
; APPLICANT: Janet A. Rice
; TITLE OF INVENTION: Synthetic Storage Proteins with Defined Structure Containing P
; NUMBER OF SEQUENCES: 113
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: E.I. du Pont de Nemours and Company
; STREET: 1007 Market Street
; CITY: Wilmington
; STATE: Delaware
; COUNTRY: USA
; ZIP: 19898
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy Disk
; COMPUTER: Macintosh
; OPERATING SYSTEM: Macintosh System, 6.0
; SOFTWARE: Microsoft Word, 4.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US92/06412
; FILING DATE: 19920807
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/743,006

```

Query Match 31.3%; Score 59.5; DB 1; Length 56;  
Best Local Similarity 44.4%; Pred. No. 0.48;

; Filing NO. 5773891  
; GENERAL INFORMATION:  
; APPLICANT: E. I. DU PONT DE NEMOURS AND

Search completed: January 30, 2004, 13:23:19  
Job time : 23 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 30, 2004, 13:18:38 ; Search time 35 Seconds  
(without alignments)  
287.544 Million cell updates/sec

Title: US-10-025-598-2

Perfect score: 190

Sequence: 1 MDWLKARVEQELQALEARCTDSNAELRAMEAKLKAETIQK 39

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- SPREMBL\_23.\*
- 1: sp\_archaea.\*
- 2: sp\_bacteria.\*
- 3: sp\_fungi.\*
- 4: sp\_human.\*
- 5: sp\_invertebrate.\*
- 6: sp\_mammal.\*
- 7: sp\_mhc.\*
- 8: sp\_organelle.\*
- 9: sp\_phase.\*
- 10: sp\_plant.\*
- 11: sp\_rodent.\*
- 12: sp\_virus.\*
- 13: sp\_vertebrate.\*
- 14: sp\_unclassified.\*
- 15: sp\_rvirus.\*
- 16: sp\_bacteriap.\*
- 17: sp\_archaeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	62	32.6	210	16	Q8XXF0
2	60.5	31.8	1549	5	Q20060
3	60	31.6	680	11	Q8R370
4	59.5	31.3	594	11	Q8CHP2
5	59	31.1	352	10	Q94GZ7
6	58.5	30.8	705	10	Q9SU64
7	58.5	30.8	717	16	Q9KTG5
8	58.5	30.8	788	11	Q8CF87
9	58.5	30.8	851	11	Q8CF88
10	58.5	30.8	909	10	Q9SU09
11	58	30.5	358	11	Q922X5
12	58	30.5	1941	5	Q26079
13	58	30.5	1950	5	Q26080
14	58	30.5	2142	10	Q9FWR3
15	57.5	30.3	122	17	Q97Z14
16	57.5	30.3	153	10	Q9M2J3

17	57.5	30.3	259	16	Q9RZ09
18	57.5	30.3	661	12	Q9PYR3
19	57.5	30.3	998	16	Q8D024
20	57.5	30.3	1052	2	Q9F291
21	57	30.0	306	16	Q8PPJ5
22	57	30.0	356	11	Q99JP6
23	57	30.0	358	11	Q92215
24	57	30.0	1052	11	Q9ERAL
25	57	30.0	1243	11	Q9ERAS
26	56.5	29.7	174	12	Q86999
27	56.5	29.7	334	16	Q9RJS9
28	56.5	29.7	836	11	Q8VDS2
29	56	29.5	211	16	Q8G4L3
30	56	29.5	227	16	Q8EG07
31	56	29.5	353	16	Q9A776
32	56	29.5	504	16	Q8PAS9
33	56	29.5	647	16	Q84627
34	56	29.5	665	11	Q91MF2
35	56	29.5	665	11	Q9DC21
36	56	29.5	842	3	Q96WP6
37	56	29.5	1078	10	Q8LI05
38	56	29.5	1945	5	Q9BLD0
39	55.5	29.2	244	3	Q74734
40	55.5	29.2	279	4	Q9ET30
41	55.5	29.2	499	4	Q9BTW3
42	55.5	29.2	507	4	Q9ET31
43	55.5	29.2	510	11	Q8KIY6
44	55.5	29.2	546	11	Q9D3K3
45	55.5	29.2	548	4	Q9ET33

ALIGNMENTS

RESULT 1

Q8XXF0	PRELIMINARY;	PRT;	210 AA.
ID	Q8XXF0		
AC	Q8XXF0;		
DT	01-MAR-2002 (TrEMBLrel. 20, Created)		
DT	01-MAR-2002 (TrEMBLrel. 20, Last sequence update)		
DT	01-OCT-2002 (TrEMBLrel. 22, Last annotation update)		
DE	Probable signal peptide protein.		
GN	RSC2164 OR RS01436.		
OS	Ralstonia solanacearum (Pseudomonas solanacearum).		
OC	Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;		
OC	Ralstoniaceae; Ralstonia.		
OX	NCBI_TaxID=305;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=GMI1000;		
RX	MEDLINE=21681879; PubMed=11823852;		
RA	Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,		
RA	Arlat M., Billault A., Brottier P., Camus J.C., Catolico L.,		
RA	Chandler M., Choise N., Claudel-Renard C., Cunnac S., Demange N.,		
RA	Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schlex T.,		
RA	Sigulier P., Thebault P., Whalen M., Wincker P., Levy M.,		
RA	Weissenbach J., Boucher C.A.;		
RT	"Genome sequence of the plant pathogen Ralstonia solanacearum.";		
RL	Nature 415:497-502(2002).		
DR	EMBL; AL64068; CAD15871.1; -		
KW	Complete proteome.		
SQ	SEQUENCE 210 AA; 23230 MW; 482BB9EF959BA875 CRC64;		
Query Match	32.6%;	Score 62;	DB 16; Length 210;
Best Local Similarity	37.1%;	Pred. No. 13;	
Matches 13;	Conservative 11;	Mismatches 11;	Indels 0; Gaps
QY	5	KARVEQELQALEARCTDSNAELRAMEAKLKAETIQK 39	
Db	131	KQALEEAQAEQRAALSQEVVRMQAQQQMLEE 165	

RESULT 2

10/025 598  
Search Notes  
2/2/04

Q8CHP2	Q8CHP2
ID	Q8CHP2
AC	Q8CHP2;



Matches 16; Conservative 8; Mismatches 11; Indels 3; Gaps 3

QY 4 LKARVEQELQA---LEARGTDSNAELPAMEAKLKAEIQ 38

RESULT 10  
Q9SU09  
ID Q9SU09 PRELIMINARY; PRT; 909 AA.



```
DR InterPro; IPR002928; Myosin_tail.
DR InterPro; IPR002017; Spectrin.
DR Pfam; PF00612; IQ; 2.
DR Pfam; PF00063; myosin_head; 1.
DR Pfam; PF02736; Myosin_N; 1.
DR Pfam; PF01576; Myosin_tail; 1.
DR PRINTS; PR00193; MYOSINHEAVY.
DR ProDom; PD000355; myosin_head; 1.
DR SMART; SM00015; IQ; 1.
DR SMART; SM00242; MYSC; 1.
DR PROSITE; PS00096; IQ; 1.
SQ SEQUENCE 1950 AA; 224184 MW; 6FE358802352C19A CRC64;

Query Match      30.5%; Score 58; DB 5; Length 1950;
Best Local Similarity 37.1%; Pred. No. 3.9e+02;
Matches 13; Conservative 8; Mismatches 14; Indels 0; Gaps 0;

QY 1 MDWLKARVEQLQALEARGTDSNAELRAMEAKLKA 35
DB 1413 LDKAKSLQLEEDLSIEVDRAVNVNQMEKKQRA 1447

RESULT 14
Q9FWR3 PRELIMINARY; PRT; 2142 AA.
AC Q9FWR3;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE F17F16.1 protein.
GN F17F16.1.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosid II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Federspiel N.A., Palm C.J., Conway A.B., Conn L., Hansen N.F.,
RA Altafi H., Nguyen M., Lam B., Southwick A., Miranda M., Brooks S.,
RA Buehler E., Chao Q., Chin C., Chiou J., Choi E., Gonzalez A.,
RA Howing B., Johnson-Hopson C., Khan S., Kim C., Koo T., Lee J.M.,
RA Lenz C., Liu A., Liu S., Mukharsky N., Pham P., Sakano H., Shinn P.,
RA Toriumi M., Vaysberg M., Yu G., Ecker J., Theologis A., Davis R.W.;
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC026237; AACG09081.1;
DR InterPro; IPR001410; DEAD.
DR SMART; SM00487; DEXDC; 1.
DR SMART; SM00487; DEXDC; 1.
SQ SEQUENCE 2142 AA; 240658 MW; B6EF6AA017A42CA2 CRC64;

Query Match      30.5%; Score 58; DB 10; Length 2142;
Best Local Similarity 31.4%; Pred. No. 4.3e+02;
Matches 11; Conservative 12; Mismatches 12; Indels 0; Gaps 0;

QY 5 KARVEQLQALEARGTDSNAELRAMEAKLKAETQK 39
DB 1451 KRKIYDLSAVQAQERKANVEMRTLKQLRKSLK 1485

RESULT 15
Q97Z14 PRELIMINARY; PRT; 122 AA.
AC Q97Z14;
DT 01-OCT-2001 (TrEMBLrel. 18, Created)
DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DT 01-OCT-2001 (TrEMBLrel. 18, Last annotation update)
DE Hypothetical protein SSO0922.
GN SSO0922.
OS Sulfolobus solfataricus.
OC Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;
OC Sulfolobus.
OX NCBI_TaxID=2287;
RN [1]
```

```
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 35092 / DSM 1617 / P2;
RX MEDLINE=21332296; PubMed=11427726;
RA She O., Singh R.K., Confalonieri F., Zivanovic Y., Allard G.,
RA Awauez M.J., Chan-Weiher C.C.-Y., Clausen I.G., Curtis B.A.,
RA De Moors A., Erauso G., Fletcher C., Gordon P.M.K.,
RA Heikamp-de Jong I., Jeffries A.C., Kozera C.J., Medina N., Peng X.,
RA Thi-Ngoc H.P., Redder P., Schenk M.E., Theriault C., Tolstrup N.,
RA Charlebois R.L., Doolittle W.F., Duguet M., Gaasterland T.,
RA Garrett R.A., Ragan M.A., Sensen C.W., Van der Oost J.;
RT "The complete genome of the crenarchaeon Sulfolobus solfataricus P2.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:7835-7840(2001).
DR EMBL; AS006712; AAK41204.1; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 122 AA; 14725 MW; 944101A92BD87F15 CRC64;

Query Match      30.3%; Score 57.5; DB 17; Length 122;
Best Local Similarity 31.6%; Pred. No. 25;
Matches 12; Conservative 11; Mismatches 8; Indels 7; Gaps 1;

QY 2 DMLKARVEQLQALEARGTDSNAELRAMEAKLKAETQK 39
DB 9 DMVKSMEKEINRME-----KEFRIESEKEIEK 39

Search completed: January 30, 2004, 13:22:11
Job time : 35 secs
```

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 30, 2004, 13:18:03 ; Search time 11 Seconds  
(without alignments)  
166.731 Million cell updates/sec

Title: US-10-025-598-2

Perfect score: 190

Sequence: 1 MDWLKARVEQLQALEARCTDSNAELRAMEAKLAEIQK 39

Scoring table:  
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_41.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match %	DB ID	Description
1	60.5	31.8	794	Q00547 mus musculus
2	60.5	31.8	1549	Q20060 caenorhabdi
3	59	31.1	316	Q03247 bos taurus
4	58.5	30.8	224	Q9h444 homo sapien
5	58.5	30.8	224	Q9d8b3 mus musculus
6	58.5	30.8	624	P40750 bacillus su
7	58	30.5	177	O8yne3 anabaena sp
8	57	30.0	1243	Q9era5 microtus ar
9	56	29.5	107	Q28995 salminki sci
10	56	29.5	574	P11516 mus musculus
11	56	29.5	665	P48678 mus musculus
12	56	29.5	665	P48679 rattus norv
13	55.5	29.2	186	P57047 haemophilus
14	55.5	29.2	577	P33240 homo sapien
15	55.5	29.2	595	P35240 homo sapien
16	55.5	29.2	596	P46662 mus musculus
17	55	28.9	298	P23529 cavia porce
18	55	28.9	664	P02545 homo sapien
19	55	28.9	864	Q02799 sus scrofa
20	54.5	28.7	222	Q9by43 homo sapien
21	54.5	28.7	854	Q9nr15 homo sapien
22	54	28.4	171	Q9acj8 thermus the
23	54	28.4	182	Q8diq8 synchococc
24	54	28.4	454	P94437 bacillus su
25	54	28.4	970	Q4222 comamonas c
26	54	28.4	2442	Q9bv73 homo sapien
27	53.5	28.2	1976	Q27991 bos taurus
28	53.5	28.2	1976	P35580 homo sapien
29	53	27.9	151	Q15007 homo sapien
30	53	27.9	1008	Q45244 caenorhabdi
31	53	27.9	1690	Q9vj65 drosophila
32	53	27.9	2349	P12270 homo sapien
33	52.5	27.6	169	P11987 methylococc

RESULT 1  
HMMR\_MOUSE  
ID HMMR\_MOUSE STANDARD; PRT; 794 AA.  
AC Q00547;  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Hyaluronan mediated motility receptor (intracellular hyaluronic acid binding protein) (Receptor for hyaluronan-mediated motility).  
GN HMMR OR IHABP OR RHAMM.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A., AND SUBCELLULAR LOCATION.  
RC TISSUE=Lung;  
RX MEDLINE=98264863; PubMed=9601097;  
RA Hofmann M., Fieber C., Asmann V., Goettlicher M., Sleeman J., Plug R., Howells N., von Stein O., Ponta H., Herrlich P.;  
RT "Identification of IHABP, a 95 kDa intracellular hyaluronate binding protein.";  
RL J. Cell Sci. 111:1673-1684(1998).  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Zhao Y., Zhang S., Turley E.;  
RT Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE OF 1-183 FROM N.A., AND TISSUE SPECIFICITY.  
RC STRAIN=129/Sv;  
RX MEDLINE=99107769; PubMed=9899313;  
RA Fieber C., Plug R., Sleeman J., Dall P., Ponta H., Hofmann M.;  
RT "Characterization of the murine gene encoding the intracellular hyaluronan receptor IHABP.";  
RL Gene 226:41-50(1999).  
RN [4]  
RP SEQUENCE OF 164-794 FROM N.A., AND ALTERNATIVE SPLICING.  
RC STRAIN=BALB/C; TISSUE=Fibroblast;  
RX MEDLINE=96011639; PubMed=7590272;  
RA Entwistle J., Zhang S., Yang B., Wong C., Li Q., Hall C.L., A J., Mowat M., Greenberg A.H., Turley E.A.;  
RT "Characterization of the murine gene encoding the hyaluronan receptor RHAMM.";  
RL Gene 163:233-238(1995).  
RN [5]  
RP SEQUENCE OF 318-794 FROM N.A., AND CHARACTERIZATION.  
RC STRAIN=BALB/C;  
RX MEDLINE=92299690; PubMed=1376732;  
RA Hardwick C., Hoare K., Owens R., Hohn H.P., Hook M., Moore D., Cripps V., Austen L., Nance D.M., Turley E.A.;  
RT "Molecular cloning of a novel hyaluronan receptor that mediates tumor cell motility.";  
RL J. Cell Biol. 117:1343-1350(1992).  
RN [6]  
RP CHARACTERIZATION.  
RX MEDLINE=94308286; PubMed=7518470;

## ALIGNMENTS

10/025,598  
Search Notes  
2/21/04

RA Hall C.L., Wang C., Lange L.A., Turley E.A.;  
RT "Hyaluronan and the hyaluronan receptor RHAMM promote focal adhesion  
RL turnover and transient tyrosine kinase activity."; J. Cell Biol. 126:575-588 (1994).  
RN [7]  
RP ERK REGULATION, AND SUBCELLULAR LOCATION.  
RX MEDLINE=98225222; PubMed=9556628;  
RA Zhang S., Chang M.C., Zylka D., Turley S., Harrison R., Turley E.A.;  
RT "The hyaluronan receptor RHAMM regulates extracellular-regulated  
RT kinase."; J. Biol. Chem. 273:11342-11348 (1998).  
RL [8]  
RN REVIEW.  
RP MEDLINE=99059494; PubMed=9845361;  
RX Hofmann M., Asemann V., Fieber C., Sleeman J.P., Moll J., Ponta H.,  
RA Hart I.R., Herrlich P.;  
RT "Problems with RHAMM: a new link between surface adhesion and  
RT oncogenesis?";  
RL Cell 95:591-592 (1998).  
CC -I- FUNCTION: INVOLVED IN CELL MOTILITY. WHEN HYALURONAN BINDS TO  
CC HMHR, THE PHOSPHORYLATION OF A NUMBER OF PROTEINS, INCLUDING THE  
CC FOCAL ADHESION KINASE OCCURS. MAY ALSO BE INVOLVED IN CELLULAR  
CC TRANSFORMATION AND METASTASIS FORMATION, AND IN REGULATING  
CC EXTRACELLULAR-REGULATED KINASE (ERK) ACTIVITY.  
CC -I- SUBUNIT: SUBUNIT OF THE HARC COMPLEX.  
CC -I- SUBCELLULAR LOCATION: CELL SURFACE AND CYTOPLASMIC.  
CC -I- ALTERNATIVE PRODUCTS:  
CC Event=Alternative splicing; Named isoforms=2;  
CC Name=RHAMM1V4;  
CC IsoId=000547-1; Sequence=Displayed;  
CC Name=RHAMM1;  
CC IsoId=000547-2; Sequence=VSP\_004287;  
CC -I- TISSUE SPECIFICITY: Ubiquitously expressed.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
DR EMBL; AF031932; AAC12655.1; -  
DR EMBL; AF079222; AAD08670.1; -  
DR EMBL; AJ005919; CAA06768.1; -  
DR EMBL; AJ005920; CAA06768.1; JOINED.  
DR EMBL; AJ005921; CAA06768.1; JOINED.  
DR EMBL; AJ005922; CAA06768.1; JOINED.  
DR EMBL; AJ005923; CAA06768.1; JOINED.  
DR EMBL; AJ005924; CAA06768.1; JOINED.  
DR EMBL; X64550; CAA45849.1; -  
DR EMBL; X64550; CAA45848.1; -  
DR PIR; JC4298; JC4298.  
DR MGD; MGI:104667; Hmnr.  
KW Hyaluronic acid; Alternative splicing; Repeat; Glycoprotein.  
FT DOMAIN 442 546 5 X 21 AA NEAR PERFECT TANDEM REPEATS.  
FT REPEAT 442 462 1.  
FT REPEAT 462 483 2.  
FT REPEAT 463 483 3.  
FT REPEAT 484 504 4.  
FT REPEAT 505 525 5.  
FT REPEAT 526 546 5.  
FT DOMAIN 719 729  
FT DOMAIN 741 750  
FT CARBOHYD 53 53  
FT CARBOHYD 134 134  
FT CARBOHYD 279 279  
FT CARBOHYD 446 446  
FT CARBOHYD 467 467  
FT CARBOHYD 488 488  
FT CARBOHYD 509 509  
FT CARBOHYD 530 530  
FT CARBOHYD 561 561  
FT CARBOHYD 601 601

FT VARSELIC 218 242 Missing (in isoform RHAMM1).  
FT CONFLICT 19 19 /FTID=VSP\_004287.  
FT CONFLICT 55 55 P -> Q (IN REF. 3).  
FT CONFLICT 71 71 S -> N (IN REF. 2).  
FT CONFLICT 89 91 K -> T (IN REF. 2).  
FT CONFLICT 94 94 EKE -> QKH (IN REF. 2).  
FT CONFLICT 94 94 A -> V (IN REF. 3).  
FT CONFLICT 540 540 S -> T (IN REF. 2).  
FT CONFLICT 668 668 E -> D (IN REF. 2).  
SQ SEQUENCE 794 AA; 91799 MW; 74DB3D236224499C CRC64;  
Query Match 31.8%; Score 60.5; DB 1; Length 794;  
Best Local Similarity 40.0%; Pred. No. 18;  
Matches 16; Conservative 9; Mismatches 8; Indels 7; Gaps 2;  
QY 7 RVEQEQLQA-LEARGT-----DSNAELRAWEAKLKAIEIOK 39  
Db 87 RUEKEIRALLQERGTDKRIQDMESELEKTEAKLNAVRE 126  
RESULT 2  
SMC4\_CAEEL  
ID\_SMC4\_CAEEL STANDARD; PRT; 1549 AA.  
AC Q20060;  
DT 15-SEP-2003 (Rel. 42, Created)  
DT 15-SEP-2003 (Rel. 42, Last sequence update)  
DT 15-SEP-2003 (Rel. 42, Last annotation update)  
DE Structural maintenance of chromosome 4 (protein smc-4).  
OS SMC-4 OR F35G12.8.  
ON Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;  
OC Rhabditidae; Peloderinae; Caenorhabditis.  
OX NCBI\_TaxID=6239;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Bristol N2;  
RA Chui C.;  
RL Submitted (OCT-1994) to the EMBL/GenBank/DBJ databases.  
RP REVISIONS.  
RA Waterston R.;  
RL Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP FUNCTION, SUBCELLULAR LOCATION, AND INTERACTION WITH MIX-1.  
RX MEDLINE=21912065; PubMed=11914278;  
RA Hagstrom K.A., Holmes V.F., Cozzarelli N.R., Meyer B.J.;  
RT "C. elegans condensin promotes mitotic chromosome architecture,  
RT centromere organization, and sister chromatid segregation during  
RT mitosis and meiosis."; Genes Dev. 16:729-742 (2002).  
RL -I- FUNCTION: Central component of the condensin complex, a complex  
CC required for conversion of the condensin complex into mitotic-like  
CC condense chromosomes. The condensin complex probably introduces  
CC positive supercoils into relaxed DNA in the presence of type I  
CC topoisomerases and converts nicked DNA into positive knotted forms  
CC in the presence of type II topoisomerases. Also involved in  
CC chromosome segregation.  
CC -I- SUBUNIT: Forms an heterodimer with mix-1. Component of the  
CC condensin complex, which contains the mix-1 and smc4 heterodimer,  
CC and probably non smc subunits that regulate the complex.  
CC -I- SUBCELLULAR LOCATION: Nuclear and cytoplasmic. In interphase  
CC cells, the majority of the condensin complex is found in the  
CC cytoplasm, while a minority of the complex is associated with  
CC chromatin. A subpopulation of the complex however remains  
CC associated with chromosome foci in interphase cells. During  
CC mitosis, most of the condensin complex is associated with the  
CC chromatin. At the onset of prophase, condensin associates with  
CC chromosome arms and to chromosome condensation. Dissociation from  
CC chromosomes is observed in late telophase.  
CC -I- DOMAIN: The flexible hinge domain, which separates the large  
CC intramolecular coiled coil regions, allows the heterodimerization  
CC with mix-1, forming a V-shaped heterodimer (by similarity).  
CC -I- SIMILARITY: Belongs to the SMC family. SMC4 subfamily.



```

RA Marsh V.L., Martin S.L., McConnachie L.J., McRay K., McMurray A.A.,
RA Milne S.A., Mistry D., Moore M.J.F., Mullikin J.C., Nickerson T.,
RA Oliver K., Parker A., Patel R., Pearce T.A.V., Peck A.I.,
RA Phillimore B.J.C.T., Prathalingam S.R., Plumb R.W., Ramsay H.,
RA Rice C.M., Ross M.T., Scott C.E., Sehra H.K., Shownkeen R., Sims S.,
RA Struebing R.L., Smith M.L., Soderlund C., Steward C.A., Sulston J.E.,
RA Swann R.M., Sycamore N., Taylor R., Tee L., Thomas D.W., Thorpe A.,
RA Tracey A., Tromans A.C., Vaudin M., Wall M., Wallis J.M.,
RA Whitehead S.L., Whittaker P., Willey D.L., Williams L., Williams S.A.,
RA Wilming L., Wray P.W., Hubbard T., Durbin R.M., Bentley D.R., Beck S.,
RA Rogers J.;
RT "The DNA sequence and comparative analysis of human chromosome 20.";
RL Nature 414:865-871(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain, Lung, and Testis;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Shenmen C.M., Schuler G.D.,
RA Klausner R.D., Collins P.S., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Altschul S.F., Zeeberg B., Moore T., Max S.I., Wang J., Hsieh F.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickinson M.C.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalhus D.E.,
RA Rodriguez A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC -1- SIMILARITY: BELONGS TO THE SNF7 FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AL050349; CAC14088.1; -
CC DR EMBL; BC033859; AAH33859.1; -
CC DR Genew; HGNC:16171; C20orf178.
CC DR InterPro; IPR005024; DUF279.
CC DR Pfam; PF03357; SNF7; 1.
SQ SEQUENCE 224 AA; 24950 MW; DB1D79D3D3803CB2F CRC64;

Query Match 30.8%; Score 58.5; DB 1; Length 224;
Best Local Similarity 36.8%; Pred. No. 8.7;
Matches 14; Conservative 11; Mismatches 12; Indels 1; Gaps 1;

Oy 2 DLKARVEQELQALAEARGT-DSNAELRAEAKLKAETQ 38
: : : : : : : : : : : : : : : : : : : : : : : : : :
Db 41 EFLEKTEQELTAAGKHGKTGNKRAALQALKRKRKYEKQ 78

RESULT 5
CTH8 MOUSE
ID CTH8 MOUSE STANDARD; PRT; 224 AA.
AC Q9D8B3; Q91VM7; Q922P1;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Protein c20orf178 homolog.
GN C20ORF178.
OS Mus musculus (Mouse).

```

```

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Small intestine;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fletschmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohetsuki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Breast tumor;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Shenmen C.M., Schuler G.D.,
RA Klausner R.D., Collins P.S., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Altschul S.F., Zeeberg B., Moore T., Max S.I., Wang J., Hsieh F.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickinson M.C.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalhus D.E.,
RA Rodriguez A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC -1- SIMILARITY: BELONGS TO THE SNF7 FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AK008205; -; NOT_ANNOTATED_CDS.
CC DR EMBL; BC006905; AAH06905.1; ALT_INIT.
CC DR EMBL; BC011429; AAH11429.1; -
CC DR MGD; MGI:1922858; 2010012F05Rik.
CC DR InterPro; IPR005024; DUF279.
CC DR Pfam; PF03357; DUF279; 1.
CC DR CONFLICT 167 167 E -> K (IN REF. 1).
SQ SEQUENCE 224 AA; 24936 MW; DB1D79C3C9ECCB2F CRC64;

Query Match 30.8%; Score 58.5; DB 1; Length 224;

```





```

CC -1- SIMILARITY: Belongs to the IF-3 family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AP003597; BAB76322.1; ALT_INIT.
CC HAMAP; MF 00080; -.
CC DR InterPro; IPR001286; IF3.
CC DR Pfam; PF00707; IF3_C; 1.
CC DR Pfam; PF05198; IF3_N; 1.
CC DR ProDom; PD002880; IF3; 1.
CC DR TIGRFAMs; TIGR00168; InfC; 1.
CC DR PROSITE; PS00938; IF3; 1.
CC KW Initiation factor; Protein biosynthesis; Complete proteome.
CC SQ SEQUENCE 177 AA; 20750 MW; 9C1585FF62590E9F CRC64;

Query Match 30.58; Score 58; DB 1; Length 177;
Best Local Similarity 36.18; Pred. No. 7.9;
Matches 13; Conservative 9; Mismatches 14; Indels 0; Gaps 0;

QY 1 MDWLKARVEQELQALEARGTDSNAELRAMEAKLKAE 36
DB 68 MDYKYEQEBQKAREARKQHTADVKEVMRYKIE 103

RESULT 8
SMC4 MICAR STANDARD; PRT; 1243 AA.
ID _SMC4 MICAR STANDARD; PRT; 1243 AA.
AC Q9ERA5; Q9ERA1;
DT 15-SEP-2003 (Rel. 42, Created)
DT 15-SEP-2003 (Rel. 42, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Structural maintenance of chromosomes 4-like 1 protein (Chromosome-
DE associated polypeptide C) (XCAP-C homolog) (Fragment).
GN SMC4L1 OR SMC4 OR CAPC.
OS Microtus arvalis (Common vole) (Field vole).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Arvicolinae;
OC Microtus.
OX NCBI_TaxID=47230;
RN [1]
RP SEQUENCE FROM N.A.
RA Pavlova S.V., Nesterova T.B., Zakian S.M.;
RT "SMC genes from common vole Microtus arvalis.";
RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: Central component of the condensin complex, a complex
CC required for conversion of interphase chromatin into mitotic-like
CC condense chromosomes. The condensin complex probably introduces
CC positive supercoils into relaxed DNA in the presence of type I
CC topoisomerases and converts nicked DNA into positive knotted forms
CC in the presence of type II topoisomerases (By similarity).
CC -1- SUBUNIT: Forms a heterodimer with SMC2L1. Component of the
CC condensin complex, which contains the SMC2L1 and SMC4L1
CC heterodimer, and three non SMC subunits that probably regulate the
CC complex: BRN1/CAPB, CNA1/CAPD2 and CAPG (By similarity).
CC -1- SUBCELLULAR LOCATION: Nuclear and cytoplasmic. In interphase
CC cells, the majority of the condensin complex is found in the
CC cytoplasm, while a minority of the complex is associated with
CC chromatin. A subpopulation of the complex however remains
CC associated with chromosome foci in interphase cells. During
CC mitosis, most of the condensin complex is associated with the
CC chromatin. At the onset of prophase, the regulatory subunits of
CC the complex are phosphorylated by CDC2, leading to condensin's
CC association with chromosome arms and to chromosome condensation.
CC Dissociation from chromosomes is observed in late telophase (By
CC similarity).
CC -1- DOMAIN: The hinge domain, which separates the large intramolecular
CC coiled coil regions, allows the heterodimerization with SMC2L1,

```

```

CC forming a V-shaped heterodimer (By similarity).
CC -1- SIMILARITY: Belongs to the SMC family. SMC4 subfamily.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AJ299713; CAC09583.1; -.
CC DR EMBL; AJ299717; CAC09587.1; -.
CC DR InterPro; IPR003439; ABC transporter.
CC DR InterPro; IPR003405; SMC_C.
CC DR InterPro; IPR003395; SMC_N.
CC DR Pfam; PF02483; SMC_C; 1.
CC DR Pfam; PF02463; SMC_N; 1.
CC KW DNA condensation; Mitosis; Cell cycle; ATP-binding; Coiled coil;
CC Nuclear protein.
CC FT NP_BIND 68 75 ATP (POTENTIAL).
CC FT DOMAIN 225 546 COILED COIL (POTENTIAL).
CC FT DOMAIN 547 720 FLEXIBLE HINGE.
CC FT DOMAIN 721 975 COILED COIL (POTENTIAL).
CC FT DOMAIN 1062 1090 COILED COIL (POTENTIAL).
CC FT DOMAIN 1146 1181 ALA/ASP-RICH (DA-BOX).
CC FT DOMAIN 21 26 POLY-PRO.
CC FT CONFLICT 241 241 N -> K (IN REF. 1; CAC09587).
CC FT CONFLICT 757 757 E -> R (IN REF. 1; CAC09587).
CC SQ SEQUENCE 1243 AA; 142341 MW; B621BD6D954DA7F9 CRC64;

Query Match 30.0%; Score 57; DB 1; Length 1243;
Best Local Similarity 39.4%; Pred. No. 69;
Matches 13; Conservative 7; Mismatches 13; Indels 0; Gaps 0;

QY 7 RVEQLQALEARGTDSNAELRAMEAKLKAEIQK 39
DB 898 RTEKEIKDTKETNDLKAELEAKAEVKK 930

RESULT 9
APE SAISC STANDARD; PRT; 107 AA.
ID _APE SAISC STANDARD; PRT; 107 AA.
AC Q28595;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Apolipoprotein E (Apo-E) (Fragment).
GN APOE.
OS Saimiri sciureus (Common squirrel monkey).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Platyrrhini; Cebidae; Saimiri.
OX NCBI_TaxID=9521;
RN [1]
RP SEQUENCE FROM N.A.
RA Morelli L., Wei L., Amorin A., McDermid J., Abec C.R., Frangione B.,
RA Walker L.C., Levy E.;
RT "Cerebrovascular amyloidosis in squirrel monkeys and rhesus monkeys:
RT apolipoprotein E genotype.";
RL FEBS Lett. 379:132-134 (1996).
CC -1- FUNCTION: APO-E MEDIATES BINDING, INTERNALIZATION, AND CATABOLISM
CC OF LIPOPROTEIN PARTICLES. IT CAN SERVE AS A LIGAND FOR THE LDL(APO
CC E/E) RECEPTOR AND FOR THE SPECIFIC APO-E RECEPTOR (CHYLOMICRON
CC REMNANT) OF HEPATIC TISSUES.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- TISSUE SPECIFICITY: Secreted in plasma.
CC -1- SIMILARITY: BELONGS TO THE APOA1 / APOA4 / APOE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its

```

use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL; U52029; AAC50442.1; --  
 HSSP; P02649; INFN.  
 InterPro; IPR000074; Apolipoprotein.  
 Pfam; PF01442; Apolipoprotein; 1.  
 Glycoprotein; Plasma; Lipid transport; HDL; VLDL; Chylomicron;  
 Heparin-binding; Repeat.  
 KW NON\_TER 1 1  
 FT DOMAIN 89 99 LDL RECEPTOR BINDING (POTENTIAL).  
 FT DOMAIN 93 96 HEPARIN-BINDING (BY SIMILARITY).  
 FT DOMAIN 11 >107 8 X 22 AA APPROXIMATE TANDEM REPEATS.  
 FT REPEAT 11 32 1.  
 FT REPEAT 33 54 2.  
 FT REPEAT 55 76 3.  
 FT REPEAT 77 98 4.  
 FT REPEAT 99 >107 5.  
 FT NON\_TER 107 107  
 SQ SEQUENCE 107 AA; FF8CED47BD18F7C CRC64;

Query Match 29.5%; Score 56; DB 1; Length 107;  
 Best Local Similarity 40.0%; Pred. No. 8;  
 Matches 14; Conservative 11; Mismatches 8; Indels 2; Gaps 2;  
 QY 5 KARVELOLEAR-GTDSNAELRAMEAKLKARIQ 38  
 Db 39 KARLSKELQAARLGADME-DVRSRLQAYRSEVQ 72

## RESULT 10

## LAMC\_MOUSE

ID\_LAMC\_MOUSE STANDARD; PRT; 574 AA.  
 AC P11516;  
 DT 01-OCT-1989 (Rel. 12, Created)  
 DT 01-FEB-1996 (Rel. 33, Last sequence update)  
 DT 15-FEB-2003 (Rel. 42, Last annotation update)  
 DE Lamins C and C2.  
 GN LMNA OR LMN1.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=89247440; PubMed=2719959;  
 RA Riedel W., Werner D.;  
 RT "Nucleotide sequence of the full-length mouse lamin C cDNA and its  
 RT deduced amino-acid sequence."  
 RL Biochim. Biophys. Acta 1008:119-122(1989).  
 RN [2]  
 RP SEQUENCE FROM N.A. (ISOFORM C).  
 RC STRAIN=ddy; TISSUE=Testis;  
 RX MEDLINE=94244715; PubMed=8187835;  
 RA Furukawa K., Inagaki H., Hotta Y.;  
 RT "Identification and cloning of an mRNA coding for a germ  
 RT cell-specific A-type lamin in mice."  
 RL Exp. Cell Res. 212:426-430(1994).  
 RN [3]  
 RP SEQUENCE OF 235-568 FROM N.A.  
 RX MEDLINE=93144345; PubMed=7916626;  
 RA Nakajima N., Sado T.;  
 RT "Nucleotide sequence of a mouse lamin A cDNA and its deduced amino  
 RT acid sequence."  
 RL Biochim. Biophys. Acta 1171:311-314(1993).  
 RN [4]  
 RP SEQUENCE OF 1-119 FROM N.A.  
 RX MEDLINE=95300954; PubMed=7781761;  
 RA Nakajima N., Abe K.;  
 RT "Genomic structure of the mouse A-type lamin gene locus encoding  
 RT somatic and germ cell-specific lamins."

FEBS Lett. 365:108-114(1995).  
 [5]  
 RN PARTIAL SEQUENCE, AND PHOSPHORYLATION.  
 RP MEDLINE=92070490; PubMed=1959608;  
 RA Eggert M., Radomski N., Tripier D., Traub P., Jost E.;  
 RT "Identification of phosphorylation sites on murine nuclear lamin C by  
 RT RP-HPLC and microsequencing."  
 RL FEBS Lett. 292:205-209(1991).  
 CC -!- FUNCTION: LAMINS ARE COMPONENTS OF THE NUCLEAR LAMINA, A FIBROUS  
 CC LAYER ON THE NUCLEOPLASMIC SIDE OF THE INNER NUCLEAR MEMBRANE,  
 CC WHICH IS THOUGHT TO PROVIDE A FRAMEWORK FOR THE NUCLEAR ENVELOPE  
 CC AND MAY ALSO INTERACT WITH CHROMATIN.  
 CC -!- SUBCELLULAR LOCATION: Nuclear.  
 CC -!- ALTERNATIVE PRODUCTS:  
 CC Event=Alternative splicing; Named isoforms=3;  
 CC Name=B; Synonyms=C;  
 CC ISOID=P11516-1; Sequence=Displayed;  
 CC Note=Isoforms A and B are present in equal amounts in the lamina  
 CC of mammals;  
 CC Name=C; Synonyms=C2;  
 CC ISOID=P11516-2; Sequence=VSP\_002471, VSP\_002472;  
 CC Name=A;  
 CC ISOID=P48678-1; Sequence=External;  
 CC Note=Isoforms A and B are present in equal amounts in the lamina  
 CC of mammals;  
 CC -!- TISSUE SPECIFICITY: Isoform C is specifically expressed in germ  
 CC cells. This specific expression and unique structure suggests a  
 CC role for this isoform in determining the organization of nuclear  
 CC and chromosomal structures during spermatogenesis.  
 CC -!- PTM: INCREASED PHOSPHORYLATION OF THE LAMINS OCCURS BEFORE  
 CC ENVELOPE DISINTEGRATION AND PROBABLY PLAYS A ROLE IN REGULATING  
 CC LAMIN ASSOCIATIONS.  
 CC -!- PTM: Phosphorylated on Ser-392, Ser-407, Ser-409 at interphase.  
 CC -!- MISCELLANEOUS: THE STRUCTURAL INTEGRITY OF THE LAMINA IS STRICTLY  
 CC CONTROLLED BY THE CELL CYCLE AS SEEN BY THE DISINTEGRATION AND  
 CC FORMATION OF THE NUCLEAR ENVELOPE IN PROPHASE AND TELOPHASE,  
 CC RESPECTIVELY.  
 CC -!- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL; X14170; CAA32372.1; --  
 EMBL; D14850; BAA03578.1; --  
 EMBL; D49733; BAA08570.1; --  
 EMBL; D49732; BAA08570.1; JOINED.  
 EMBL; D49734; BAA08570.1; JOINED.  
 EMBL; D49735; BAA08570.1; JOINED.  
 EMBL; D49733; BAA08571.1; --  
 EMBL; D49736; BAA08571.1; JOINED.  
 EMBL; D49734; BAA08571.1; JOINED.  
 EMBL; D49735; BAA08571.1; JOINED.  
 PIR; I53414; I53414.  
 PIR; S04333; S04333.  
 MGD; MGI:96794; Lmna.  
 GO; GO:0005638; C:lamin filament; IDA.  
 InterPro; IPR001664; IF.  
 InterPro; IPR001322; IF tail.  
 Pfam; PF00038; filament\_1.  
 Pfam; PF00932; IF\_tail; 1.  
 PROSITE; PS00226; IF; 1.  
 KW Intermediate filament; Coiled coil; Nuclear protein; Phosphorylation;  
 KW Alternative splicing.  
 FT DOMAIN 1 33 HEAD.  
 FT DOMAIN 34 383 ROD.  
 FT DOMAIN 384 574 TAIL.  
 FT DOMAIN 34 70 COIL 1A.  
 FT DOMAIN 71 80 LINKER 1.

```

FT DOMAIN      81 218 COIL 1B.
FT DOMAIN      219 242 LINKER 2.
FT DOMAIN      243 383 COIL 2.
FT SITE        325 325 STUTTER (BY SIMILARITY).
FT SITE        266 266 HEPTAD CHANGE OF PHASE.
FT SITE        330 330 HEPTAD CHANGE OF PHASE.
FT SITE        417 422 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT MOD_RES     392 392 PHOSPHORYLATION (BY CDC2) (PARTIAL).
FT MOD_RES     407 407 PHOSPHORYLATION (PARTIAL).
FT MOD_RES     409 409 PHOSPHORYLATION (PARTIAL).
FT VARSPLIC    1 112 Missing (in isoform C).
FT VARSPLIC    113 118 /FTID=VSP_002471.
FT VARSPLIC    118 118 FELKA -> MGNREG (in isoform C).
FT CONFLICT    118 119 /FTID=VSP_002472.
FT CONFLICT    401 401 AR -> VC (IN REF. 1).
FT CONFLICT    439 440 R -> P (IN REF. 1).
FT CONFLICT    439 440 RV -> WL (IN REF. 1).
SQ SEQUENCE    574 AA; 65446 MW; A736DF1CCEDB65BE CRC64;

Query Match      29.5%; Score 56; DB 1; Length 574;
Best Local Similarity 37.9%; Pred. No. 42;
Matches 11; Conservative 8; Mismatches 10; Indels 0; Gaps 0;

QY 6 ARVEQELQALEARGTDSNAELRAMEAKLK 34
Db 107 SKVREPFKELKARNTKKGDLAAQARLK 135

RESULT 11
LAMA MOUSE
ID LAMA MOUSE STANDARD; PRT; 665 AA.
AC P48678; P97859;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Lamin A.
GN MNNA OR LMNI.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP MEDLINE=95300954; PubMed=7781761;
RA Nakajima N., Abe K.;
RT "Genomic structure of the mouse A-type lamin gene locus encoding
RL somatic and germ cell-specific lamins.";
RL FEBS Lett. 365:108-114 (1995).
RN [2]
RP SEQUENCE OF 1-568 FROM N.A.
RX MEDLINE=89247440; PubMed=2719959;
RA Riedel W., Werner D.;
RT "Nucleotide sequence of the full-length mouse lamin C cDNA and its
RL deduced amino-acid sequence.";
RL Biochim. Biophys. Acta 1008:119-122 (1989).
RN [3]
RP SEQUENCE OF 235-665 FROM N.A.
RX MEDLINE=93144345; PubMed=7916626;
RA Nakajima N., Sado T.;
RT "Nucleotide sequence of a mouse lamin A cDNA and its deduced amino
RL acid sequence.";
RL Biochim. Biophys. Acta 1171:311-314 (1993).
CC -1- FUNCTION: LAMINS ARE COMPONENTS OF THE NUCLEAR LAMINA, A FIBROUS
CC LAYER ON THE NUCLEOPLASMIC SIDE OF THE INNER NUCLEAR MEMBRANE,
CC WHICH IS THOUGHT TO PROVIDE A FRAMEWORK FOR THE NUCLEAR ENVELOPE
CC AND MAY ALSO INTERACT WITH CHROMATIN.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- ALTERNATIVE PRODUCTS:
CC Event-Alternative splicing; Named isoforms=3;
CC Name=A;
CC IsoId=P48678-1; Sequence=Displayed;
CC Note=isoforms A and C are present in equal amounts in the
CC lamina of mammals;

```

```

CC Name=B;
CC IsoId=P11516-1; Sequence=External;
CC Name=C;
CC IsoId=P11516-2; Sequence=External;
CC Note=isoforms A and C are present in equal amounts in the
CC lamina of mammals;
CC -1- PTM: INCREASED PHOSPHORYLATION OF THE LAMINS OCCURS BEFORE
CC ENVELOPE DISINTEGRATION AND PROBABLY PLAYS A ROLE IN REGULATING
CC LAMIN ASSOCIATIONS.
CC -1- MISCELLANEOUS: THE STRUCTURAL INTEGRITY OF THE LAMINA IS STRICTLY
CC CONTROLLED BY THE CELL CYCLE, AS SEEN BY THE DISINTEGRATION AND
CC FORMATION OF THE NUCLEAR ENVELOPE IN PROPHASE AND TELOPHASE,
CC RESPECTIVELY.
CC -1- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.
CC -1- CAUTION: REF.1 SEQUENCE DIFFERS FROM THAT SHOWN FROM POSITION 582
CC ONWARD AND IS LONGER (702 AA) DUE TO A FRAMESHIFT.
CC
CC THIS SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; D49733; BAA08569.1; -
CC EMBL; D49732; BAA08569.1; JOINED.
CC EMBL; D49734; BAA08569.1; JOINED.
CC EMBL; D49735; BAA08569.1; JOINED.
CC EMBL; D13181; BAA02476.1; -
CC GGI; MGI:96794; Lmna.
CC GO; GO:0005638; C:lamin filament; IDA.
CC InterPro; IPR001664; IF.
CC InterPro; IPR001322; IF_tail.
CC Pfam; PF00038; filament; 1.
CC Pfam; PF00932; IF_tail; 1.
CC PROSITE; PS00226; IF; 1.
CC KW Intermediate filament; Coiled coil; Nuclear protein; Phosphorylation;
CC Alternative splicing.
FT DOMAIN      1 33 HEAD.
FT DOMAIN      34 383 ROD.
FT DOMAIN      384 665 TAIL.
FT DOMAIN      34 70 COIL 1A.
FT DOMAIN      71 80 LINKER 1.
FT DOMAIN      81 218 COIL 1B.
FT DOMAIN      219 242 LINKER 2.
FT DOMAIN      243 383 COIL 2.
FT SITE        325 325 STUTTER (BY SIMILARITY).
FT DOMAIN      417 422 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT LIPID        662 662 FARNESYL (BY SIMILARITY).
FT CONFLICT    118 119 AR -> VC (IN REF. 2).
FT CONFLICT    401 401 R -> P (IN REF. 2).
FT CONFLICT    439 440 RV -> WL (IN REF. 2).
FT CONFLICT    623 623 A -> V (IN REF. 1).
SQ SEQUENCE    665 AA; 74209 MW; 5434ED6D0A04CB15 CRC64;

Query Match      29.5%; Score 56; DB 1; Length 665;
Best Local Similarity 37.9%; Pred. No. 48;
Matches 11; Conservative 8; Mismatches 10; Indels 0; Gaps 0;

QY 6 ARVEQELQALEARGTDSNAELRAMEAKLK 34
Db 107 SKVREPFKELKARNTKKGDLAAQARLK 135

RESULT 12
LAMA RAT
ID LAMA RAT STANDARD; PRT; 665 AA.
AC P48679;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Lamin A.

```

```

GN LNVA OR LMNI.
OS Rattus norvegicus (Rat).
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OC NCBI_TaxID=10116;
RN [1]
RN SEQUENCE FROM N.A.
RP MEDLINE=93050186; PubMed=1426247;
RX Ozaki T., Sakiyama S.;
RA Jonnalagadda V.S., Parnaik V.K.;
RA Submitted (NOV-1993) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: LAMINS ARE COMPONENTS OF THE NUCLEAR LAMINA, A FIBROUS
CC LAYER ON THE NUCLEOPLASMIC SIDE OF THE INNER NUCLEAR MEMBRANE,
CC WHICH IS THOUGHT TO PROVIDE A FRAMEWORK FOR THE NUCLEAR ENVELOPE,
CC AND MAY ALSO INTERACT WITH CHROMATIN.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=Lamin A;
CC IsoId=P48679-1; Sequence=Displayed;
CC Note=Isoforms A and C are present in equal amounts in the lamina
CC of mammals;
CC Name=Lamin C;
CC IsoId=P48679-2; Sequence=Not described;
CC Note=Isoforms A and C are present in equal amounts in the lamina
CC of mammals;
CC -1- PTM: INCREASED PHOSPHORYLATION OF THE LAMINS OCCURS BEFORE
CC ENVELOPE DISINTEGRATION AND PROBABLY PLAYS A ROLE IN REGULATING
CC LAMIN ASSOCIATIONS.
CC -1- MISCELLANEOUS: THE STRUCTURAL INTEGRITY OF THE LAMINA IS STRICTLY
CC CONTROLLED BY THE CELL CYCLE, AS SEEN BY THE DISINTEGRATION AND
CC FORMATION OF THE NUCLEAR ENVELOPE IN PROPHASE AND TELOPHASE,
CC RESPECTIVELY.
CC -1- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.
CC -1- CAUTION: REF.1 SEQUENCE DIFFERS FROM THAT SHOWN IN THE N- AND C-
CC TERMINAL DUE TO FRAMESHIFTS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR ENBL; X66870; CAA47342.1; ALT_FRAME.
DR ENBL; X76297; CAA53945.1; -.
DR InterPro; IPR001664; IF.
DR InterPro; IPR001322; IF_tail.
DR Pfam; PF00038; filament; 1.
DR Pfam; PF00932; IF_tail; 1.
DR PROSITE; PS00226; IF; 1.
KW Intermediate filament; Coiled coil; Nuclear protein; Lipoprotein;
KW Prenylation; Phosphorylation; Alternative splicing.
FT DOMAIN 1 33 HEAD.
FT FT 34 383 ROD.
FT FT 384 665 TAIL.
FT FT 34 70 COIL 1A.
FT FT 71 80 LINKER 1.
FT FT 81 218 COIL 1B.
FT FT 219 242 LINKER 2.
FT FT 243 383 COIL 2.
FT FT 325 325 STUTTER (BY SIMILARITY).
FT FT 417 422 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT FT 662 662 FARNESYL (BY SIMILARITY).
FT FT 470 470 K -> R (IN REF. 1).
FT FT 524 524 T -> S (IN REF. 1).
FT FT CONFLICT 524 524

```



- RL Cell 72:791-800(1993).  
RN [2]  
RP SEQUENCE FROM N.A. (ISOFORM 1).  
RX MEDLINE=93281181; PubMed=9379998;  
RA Rouleau G.A., Merel P., Lutchman M., Sanson M., Zucman J.,  
RA Marineau C., Hoang-Xuan K., Demczuk S., Demare C., Plougastel B.,  
RA Pulst S., Lenoir G., Bijlma E., Fashold R., Dumanski J., de Jong P.,  
RA Parry D., Eldridge R., Aurias A., Delattre O., Thomas G.;  
RT "Alteration in a new gene encoding a putative membrane-organizing  
RT protein causes neuro-fibromatosis type 2.";  
RL Nature 363:515-521(1993).  
RN [3]  
RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).  
RX MEDLINE=99036680; PubMed=9817927;  
RA Zucman-Rossi J., Legoux P., Der Sarkissian H., Cheret G., Sor F.,  
RA Bernardi A., Cases L., Giraud S., Lenoir G., Thomas G.;  
RT "NF2 gene in neurofibromatosis type 2 patients.";  
RL Hum. Mol. Genet. 7:2095-2101(1998).  
RN [4]  
RP SEQUENCE FROM N.A. (ISOFORMS 7; 9 AND 10), AND SUBCELLULAR LOCATION.  
RX MEDLINE=99330566; PubMed=10401006;  
RA Schucker B., Tang Y., Kressel M.;  
RT "Novel alternatively spliced isoforms of the neurofibromatosis type 2  
RT tumor suppressor are targeted to the nucleus and cytoplasmic  
RT granules.";  
RL Hum. Mol. Genet. 8:1561-1570(1999).  
RN [5]  
RP SEQUENCE FROM N.A. (ISOFORMS 1; 3; 4; 5; 6 AND 8).  
RX MEDLINE=21696156; PubMed=11827459;  
RA Chang L.-S., Akhmetieva E.M., Wu Y., Zhu L., Welling D.B.;  
RT "Multiple transcription initiation sites, alternative splicing, and  
RT differential polyadenylation contribute to the complexity of human  
RT neurofibromatosis 2 transcripts.";  
RL Genomics 79:63-76(2002).  
RN [6]  
RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 4).  
RC TISSUE=Lung, and Skin;  
RX MEDLINE=22388257; PubMed=12477932;  
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
RA Klausner R.D., Collins F.S., Wagner L., Shenman C.M., Schuler G.D.,  
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heist F.,  
RA Datchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,  
RA Raha S.S., Loquellano N.A., Peters G.J., Abranson R.D., Mullaby S.J.,  
RA Boak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
RA Vallalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,  
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalley D.E.,  
RA Schnerch A., Schein J.B., Jones S.J.M., Marra M.A.;  
RT "Generation and initial analysis of more than 15,000 full-length  
RT human and mouse cDNA sequences.";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
RN [7]  
RP REVIEW.  
RA Marineau C., Merel P., Rouleau G.A., Thomas G.;  
RT "The gene of neurofibromatosis type 2.";  
RL Medecine/Sciences 11:35-42(1995).  
RN [8]  
RP VARIANT NF2 TYR-220.  
RX MEDLINE=94047514; PubMed=8230593;  
RA Maccollin M.M., Mohney T., Trofatter J.A., Wertelecki W., Ramesh V.,  
RA Gusella J.F.;  
RT "DNA diagnosis of neurofibromatosis 2. Altered coding sequence of the  
RT merlin tumor suppressor in an extended pedigree.";  
RL JAMA 270:2316-2320(1993).  
RN [9]  
RP VARIANT NF2 PHE-96 DEL.
- RX MEDLINE=94311312; PubMed=7913580;  
RA Maccollin M.M., Ramesh V., Jacoby L.B., Louis D.N., Rubio M.-P.,  
RA Pulaski K., Trofatter J.A., Short M.P., Bove C., Eldridge R.,  
RA Parry D.M., Gusella J.F.;  
RT "Mutational analysis of patients with neurofibromatosis 2.";  
RL Am. J. Hum. Genet. 55:314-320(1994).  
RN [10]  
RP VARIANT ARG-46.  
RX MEDLINE=94272478; PubMed=8004107;  
RA Irving R.M., Moffat D.A., Hardy D.G., Barton D.E., Xuereb J.H.,  
RA Maher E.R.;  
RT "Somatic NF2 gene mutations in familial and non-familial vestibular  
RT schwannoma.";  
RL Hum. Mol. Genet. 3:347-350(1994).  
RN [11]  
RP VARIANTS MET-219 AND CYS-418.  
RX MEDLINE=94282032; PubMed=8012353;  
RA Jacoby L.B., Maccollin M.M., Louis D.N., Mohney T., Rubio M.-P.,  
RA Pulaski K., Trofatter J.A., Kley N., Seizinger B.R., Ramesh V.,  
RA Gusella J.F.;  
RT "Exon scanning for mutation of the NF2 gene in schwannomas.";  
RL Hum. Mol. Genet. 3:413-419(1994).  
RN [12]  
RP VARIANTS NF2 SER-62; GLY-106 AND MET-352.  
RX MEDLINE=94362682; PubMed=8081368;  
RA Bourn D., Carter S.A., Mason S., Gareth D., Evans R., Strachan T.;  
RT "Germline mutations in the neurofibromatosis type 2 tumour suppressor  
RT gene.";  
RL Hum. Mol. Genet. 3:813-816(1994).  
RN [13]  
RP VARIANT GLY-79.  
RX MEDLINE=95038740; PubMed=7951231;  
RA Sainz J., Huynh D.P., Figueroa K., Ragge N.K., Baser M.E., Pulst S.M.;  
RT "Mutations of the neurofibromatosis type 2 gene and lack of the gene  
RT product in vestibular schwannomas.";  
RL Hum. Mol. Genet. 3:885-891(1994).  
RN [14]  
RP VARIANTS PHE-273 AND ILE-364.  
RX MEDLINE=94214500; PubMed=8162073;  
RA Bianchi A.B., Hara T., Ramesh V., Gao J., Klein Szanto A.J., Morin F.,  
RA Bianchi A.G., Trofatter J.A., Gusella J.F., Seizinger B.R., Kley N.;  
RT "Mutations in transcript isoforms of the neurofibromatosis 2 gene in  
RT multiple human tumour types.";  
RL Nat. Genet. 6:185-192(1994).  
RN [15]  
RP VARIANT NF2 PRO-535.  
RX MEDLINE=95395825; PubMed=7666400;  
RA Evans D.G.R., Bourn D., Wallace A., Ramsden R.T., Mitchell J.D.,  
RA Strachan T.;  
RT "Diagnostic issues in a family with late onset type 2  
RT neurofibromatosis.";  
RL J. Med. Genet. 32:470-474(1995).  
RN [16]  
RP VARIANT NF2 PRO-538.  
RX MEDLINE=96163499; PubMed=8566958;  
RA Kluewe L., Mautner V.-F.;  
RT "A missense mutation in the NF2 gene results in moderate and mild  
RT clinical phenotypes of neurofibromatosis type 2.";  
RL Hum. Genet. 97:224-227(1996).  
RN [17]  
RP VARIANTS PHE-96 DEL; ILE-117; PHE-119 DEL; 122-VAL--GLU-129 DEL AND  
RP PHE-339.  
RX MEDLINE=96209910; PubMed=8655144;  
RA de Vitis L.R., Tedde A., Vitelli F., Ammannati F., Mennonna P.,  
RA Bigozzi U., Montali E., Papi L.;  
RT "Screening for mutations in the neurofibromatosis type 2 (NF2) gene  
RT in sporadic meningiomas.";  
RL Hum. Genet. 97:632-637(1996).  
RN [18]  
RP VARIANT NF2 ARG-234.  
RX MEDLINE=99192373; PubMed=10090912;  
RA Baser M.E., Kluewe L., Mautner V.-F.;  
RT "Germ-line NF2 mutations and disease severity in neurofibromatosis

RT type 2 patients with retinal abnormalities.";

RL Am. J. Hum. Genet. 64:1230-1233(1999).

CC -!- FUNCTION: Probably acts as a membrane stabilizing protein.

CC -!- SUBCELLULAR LOCATION: In a fibroblastic cell line, isoforms 1 and

CC 10 are found homogenously distributed over the entire cell, with a

CC particularly strong staining in ruffling membranes and filopodia.

CC Isoform 10 is also found in the nucleus. Isoforms 7, 9 and 10 are

CC observed in cytoplasmic granules concentrated in a perinuclear

CC location. Isoforms 7 and 9 are absent from ruffling membranes and

CC filopodia.

CC -!- ALTERNATIVE PRODUCTS:

CC Event=Alternative splicing; Named isoforms=10;

CC Name=1; Synonyms=1;

CC IsoId=P35240-1; Sequence=Displayed;

CC Name=2;

CC IsoId=P35240-2; Sequence=VSP\_000492;

CC Name=3; Synonyms=11;

CC IsoId=P35240-3; Sequence=VSP\_007050, VSP\_007051;

CC Name=4; Synonyms=delete/3;

CC IsoId=P35240-4; Sequence=VSP\_007041, VSP\_007050, VSP\_007051;

CC Name=5; Synonyms=delete3;

CC IsoId=P35240-5; Sequence=VSP\_007042, VSP\_007050, VSP\_007051;

CC Name=6; Synonyms=delete2;

CC IsoId=P35240-6; Sequence=VSP\_007040, VSP\_007050, VSP\_007051;

CC Name=7; Synonyms=MER150;

CC IsoId=P35240-7; Sequence=VSP\_007045, VSP\_007046;

CC Name=8;

CC IsoId=P35240-8; Sequence=VSP\_007048, VSP\_007050, VSP\_007051;

CC Name=9; Synonyms=MER162;

CC IsoId=P35240-9; Sequence=VSP\_007044;

CC Name=10; Synonyms=MER151;

Query Match 29.2%; Score 55.5; DB 1; Length 595;

Best Local Similarity 42.4%; Pred.No. 49;

Matches 14; Conservative 6; Mismatches 12; Indels 1; Gaps 1;

QY 6 ARVEQELQALEARGTDSNAELAMEAK-LKAEI 37

DB 403 AEAQEMQRIKATAIRTEEEKRLMEQKVLAEV 435

Search completed: January 30, 2004, 13:21:24

Job time : 12 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 30, 2004, 13:19:19 ; Search time 21 Seconds  
(without alignments)  
178.599 Million cell updates/sec

Title: US-10-025-598-2

Perfect score: 190

Sequence: 1 MDWLKARVQELQALRGTDTSNAELRAMEAKLKAEIQK 39

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR\_76:\*

1: pir1:\*

2: pir2:\*

3: pir3:\*

4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	60.5	31.8	1549	2 T21809	hypothetical prote
2	59	31.1	316	2 JC6549	apolipoprotein E p
3	59	31.1	316	2 S26478	apolipoprotein E
4	58.5	30.8	624	2 A55220	penicillin-binding
5	58.5	30.8	705	2 T06692	hypothetical prote
6	58.5	30.8	737	2 G82262	probable exopolysa
7	58.5	30.8	909	2 T06635	hypothetical prote
8	58	30.5	204	2 AG2383	translation initia
9	58	30.5	2142	2 D86303	F17F16.1 protein -
10	57.5	30.3	122	2 E90243	hypothetical prote
11	57.5	30.3	153	2 T45983	hypothetical prote
12	57.5	30.3	259	2 C75610	hypothetical prote
13	56.5	29.7	174	2 S49245	19K protein - soil
14	56	29.5	353	2 E87478	GGDEF family prote
15	56	29.5	647	2 G71490	hypothetical prote
16	56	29.5	665	2 S27267	lamin A - rat
17	55.5	29.2	192	2 I48683	schwannomin - mous
18	55.5	29.2	244	2 T39637	hypothetical prote
19	55.5	29.2	577	2 A40220	cleavage stimulat
20	55.5	29.2	591	2 T54368	merlin protein - m
21	55.5	29.2	595	2 S33809	neurofibromin 2 -
22	55.5	29.2	596	2 I68664	merlin - mouse
23	55	28.9	278	2 S29615	whiG protein - Str
24	55	28.9	298	2 S12635	apolipoprotein E p
25	55	28.9	421	2 D82500	hypothetical prote
26	55	28.9	572	1 VEHUUC	lamin C - human
27	55	28.9	629	2 T44607	hypothetical prote
28	55	28.9	664	1 VEHULA	lamin A - human
29	54.5	28.7	845	2 T00071	hypothetical prote

30	54	28.4	329	2 JC5566	apolipoprotein E p
31	54	28.4	345	2 D91201	type III secretion
32	54	28.4	345	2 H86047	escu [imported] -
33	54	28.4	454	2 A69763	homoserine dehydro
34	54	28.4	501	2 F87550	hypothetical prote
35	54	28.4	545	2 F83280	probable chemotaxi
36	54	28.4	867	2 T14777	hypothetical prote
37	54	28.4	970	2 A41290	probable transposa
38	54	28.4	1091	2 T34107	hypothetical prote
39	54	28.4	2442	2 T08621	centrosome associa
40	53.5	28.2	188	2 D84468	hypothetical prote
41	53.5	28.2	511	2 T15629	hypothetical prote
42	53.5	28.2	627	2 D84726	hypothetical prote
43	53.5	28.2	758	2 F72363	hypothetical prote
44	53.5	28.2	934	2 AG1889	WD-40 repeat prote
45	53.5	28.2	1410	1 A57013	early endosome ant

ALIGNMENTS

RESULT 1

T21809

hypothetical protein F35G12.8 - Caenorhabditis elegans

C;Species: Caenorhabditis elegans

C;Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 02-Jun-2000

C;Accession: T21809

R;Chui, C.

submitted to the EMBL Data Library, October 1994

A;Reference number: Z19473

A;Accession: T21809

A;Status: preliminary; translated from GB/EMBL/DDBJ

A;Molecule type: DNA

A;Residues: 1-1549 <WIL>

A;Cross-references: EMBL:Z46242; PIDN:CAA86336.1; GSPDB:GN00021; CESP:F35G12.8

C;Experimental source: Clone F35G12

C;Genetics:

A;Gene: CESP:F35G12.8

A;Map position: 3

C;Introns: 50/3; 72/3; 177/3; 235/3; 333/2; 581/3; 667/3; 812/3; 882/1; 1050/3; 1223/3; 1

C;Superfamily: chromosome segregation protein SMC1

Query Match 31.8%; Score 60.5; DB 2; Length 1549;

Best Local Similarity 36.1%; Pred. No. 52;

Matches 13; Conservative 10; Mismatches 10; Indels 3; Gaps 1;

Qy 7 RVPEQLQALEARG---TDSNAELRAMEAKLKAEIQK 39

Db 990 RFEKQIKLQTKGEEMTKQSELSSAAETKLEGELKK 1025

RESULT 2

JC6549

apolipoprotein E precursor - sheep

C;Species: Ovis sp. (sheep)

C;Date: 05-Dec-1998 #sequence\_revision 05-Dec-1998 #text\_change 07-May-1999

C;Accession: JC6549

R;Komatsu, Y.; Horiuchi, M.; Ishiguro, N.; Matsui, T.; Shinagawa, M.

Gene 208, 131-138, 1998

A;Title: Characterization of the sheep apolipoprotein E (ApoE) gene and allelic variation

A;Reference number: JC6549; MUID:98201606; PMID:9524247

A;Accession: JC6549

A;Molecule type: mRNA

A;Residues: 1-316 <KOM>

C;Genetics:

A;Gene: apoE

A;Introns: 15/1; 78/2

C;Superfamily: apolipoprotein A-I

C;Keywords: lipoprotein

F;1-18/Domain: signal sequence #status predicted <SIG>

F;19-316/Product: apolipoprotein E #status experimental <MAT>

Query Match 31.1%; Score 59; DB 2; Length 316;

10/025, 598  
Search Notes  
2/21/04

A;Start codon: GTG  
C;Superfamily: penicillin-binding protein 1B

Query Match            30.8%; Score 58.5; DB 2; Length 624;  
Best Local Similarity 33.3%; Pred. No. 35;  
Matches 13; Conservative 14; Mismatches 11; Indels 1; Gaps 1;

QY     2 DMLKRVQLQEALRGATSDSNEL-RAMEAKLKAEIQ 39  
     |::|::||:|::||:~::~||:~::~||:~::~||:  
DB     226 DYTKRQRLLGLKGADGVITDKLKKAVQIKILDVEK 264

RESULT 5

T06682  
hypothetical protein T17F15.120 - Arabidopsis thaliana  
C;Species: Arabidopsis thaliana (mouse-ear cress)  
C;Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 22-Oct-1999  
C;Accession: T06682  
R;Quekter, F.; Choasne, N.; Robert, C.; Brotier, P.; Wincker, P.; Cattolico, L.; Artigues,  
A;Reference number: Z15793  
A;Accession: T06682  
A;Molecule type: DNA  
A;Residues: 1-705 <QUE>  
A;Cross-references: EMBL:AL049658; GSPDB:GN00061; ATSP:T17F15.120  
A;Experimental source: cultivar Columbia; BAC clone T17F15  
C;Genetics:  
A;Gene: ATPS-T17F15.120  
A;Map position: 3  
A;Introns: 160/3; 232/3; 338/2; 375/3; 454/3; 586/2

Query Match            30.8%; Score 58.5; DB 2; Length 705;  
Best Local Similarity 32.7%; Pred. No. 40;  
Matches 16; Conservative 8; Mismatches 14; Indels 11; Gaps 2;

QY     2 DMWKAR-VQEQLQA-----LEARGTDSNAELRAMEAKLKAEIQ 39  
     |:|:~::~||:~::~||:~::~||:~::~||:~::~||:~::~||:  
DB     397 EWHRHRLDPDELQERVRFRVQVKWLTTRGVDEEAILRALPLDLRRQIOR 445

RESULT 6

G82262  
probable exopolysaccharide biosynthesis protein VC0937 [imported] - Vibrio cholerae (strain  
C;Species: Vibrio cholerae  
C;Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 02-Feb-2001  
C;Accession: G82262  
R;Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.J.;  
Clayton, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers, P.  
l., R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.  
Nature 406, 477-483, 2000  
A;title: DNA Sequence of both chromosomes of the choleta pathogen Vibrio cholerae.  
A;reference number: A82035; MUID:20406833; PMID:10952301  
A;Accession: G82262  
A>Status: preliminary  
A:Molecule type: DNA  
A;Residues: 1-737 <HEI>  
A;Cross-references: GB:AE004176; GB:AE003852; NID:g9655385; PIDN:AAF94099.1; GSPDB:GN0011  
A;Experimental source: serogroup O1; strain NI6961; biotype El Tor  
C;Genetics:  
A;Gene: VC0937  
A;Map position: 1

Query Match            30.8%; Score 58.5; DB 2; Length 737;  
Best Local Similarity 42.1%; Pred. No. 42;  
Matches 16; Conservative 8; Mismatches 11; Indels 3; Gaps 1;

QY     4 LKARVQEQLQA---LEARGTDSNAELRAMEAKLKAEIQ 38  
     ||:~::~||:~::~||:~::~||:~::~||:~::~||:~::~||:  
DB     203 LEAKVELKANTWLNTMRMEELRTQLRESEAKLQAFLO 240

RESULT 7

T06635

A:Reference number: A86141; MUID:21016719; PMID:11130712  
A:Accession: D86303  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-2142 <STO>  
A:Cross-references: GB:AE005172; NID:g9954728; PIDN:AAG09081.1; GSPDB:GN00141  
C:Genetics:  
A:Map position: 1

Query Match            30.5%; Score 58; DB 2; Length 2142;  
Best Local Similarity 31.4%; Pred.No.1.4e+02;

Matches 11; Conservative 12; Mismatches 12; Indels 0; Gaps

Oy 5 KARVEGELQALEARGTDSNAELRAMEAKUKAEIQK 39  
     |::|||::|::|::|::|::|::|  
Db 1451 KRKIYKDLSAVQAQRKANVEMRTLKQLRKSLIK 1485

RESULT 10  
E90243  
hypothetical protein SSO0922 [imported] - Sulfolobus solfataricus  
C:Species: Sulfolobus solfataricus  
C>Date: 24-May-2001 #sequence\_revision 24-May-2001 #text\_change 24-May-2001  
C:Accession: E90243  
F:She, Q.; Singh, R.K.; Confalonieri, F.; Zivanovic, Y.; Allard, G.; Awayez,  
Jong, I.; Jeffries, A.C.; Kozera, C.J.; Medina, N.; Peng, X.; Thi-Ngoc, H.P.  
arrett, R.A.; Ragan, M.A.; Sensen, C.W.; Van der Oost, J.  
submitted to GenBank, April 2001  
A>Description: Sulfolobus solfataricus complete genome.  
A:Reference number: A99139  
A:Accession: E90243  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-122 <XUR>  
A:Cross-references: GB:AE006641; NID:gl3814104; PIDN:AAK41204.1; GSPDB:GN0015  
C:Genetics:  
A:Gene: SSO0922

Query Match            30.3%; Score 57.5; DB 2; Length 122;  
Best Local Similarity 31.6%; Pred.No.8.6;

Matches 12; Conservative 11; Mismatches 8; Indels 7; Gaps

Oy 2 DWLKAARVEGLQALREARTGSNAELRAMERAKUEIQQ 39  
     |::|||::|::|::|::|::|::|  
Db 9 DMVKSMIEKEINRM-----KEFRRIESEIKEIEK 39

RESULT 11  
T45983  
hypothetical protein F9D24.140 - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C>Date: 04-Feb-2000 #sequence\_revision 04-Feb-2000 #text\_change 04-Feb-2000  
C:Accession: T45983  
R:D'Angelo, M.; Vezzi, A.; Modesto, D.; Pigazzi, M.; Valle, G.; Mwewa, H.W.;  
submitted to The Protein Sequence Database, January 2000  
A:Reference number: Z23011  
A:Accession: T45983  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-153 <DN>  
A:Cross-references: EMBL:AlU37081  
A:Experimental source: cultivar Columbia; BAC clone F9D24  
C:Genetics:  
A:Map position: 3  
A:Introns: 19/3  
A>Note: F9D24.140

Query Match            30.3%; Score 57.5; DB 2; Length 153;  
Best Local Similarity 32.7%; Pred.No.11;

Matches 16; Conservative 9; Mismatches 11; Indels 13; Gaps

Oy 1 MDMLKARVEQ---ELOALREAG-----TDSNAELRAMERAKUKAE 36  
     |||||||::|::|::|::|::|::|::|:

Db 88 LDWLKNRLESLCKVKKKMEARGARMRELDGMIVQRRVLWALTELNKE 136

## RESULT 12

C75610

hypothetical protein - Deinococcus radiodurans (strain R1)

C;Species: Deinococcus radiodurans

C;Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 28-Jul-2000

C;Accession: C75610

R;White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.; M.; Shen, M.; Vamathevan, J.U.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Ma S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.  
Science 286, 1571-1577, 1999

A;Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.

A;Reference number: A75250; MUID:20036896; PMID:10567266

A;Accession: C75610

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-259 <WHI>

A;Cross-references: GB:AE001862; GB:AE001825; NID:g6460468; PIDN:AAF12313.1; PID:g646060

A;Experimental source: strain R1

C;Genetics:

A;Gene: DRA0144

A;Map position: 2

C;Superfamily: Deinococcus radiodurans hypothetical protein DRA0144

Query Match 30.3%; Score 57.5; DB 2; Length 259;  
Best Local Similarity 40.9%; Pred. No. 19;  
Matches 18; Conservative 6; Mismatches 9; Indels 11; Gaps 2;

QY 7 RVEQELQALE-----ARGTD--SNAEIRAMEAKLKAEIQK 39

Db 207 RVQENQALTFWDEYKTKAPVDNDSAEIRASQKQLQFLQK 250

## RESULT 13

S49245

19K protein - soil-borne wheat mosaic virus (isolate OK1-1)

C;Species: soil-borne wheat mosaic virus, SBMWV

A;Variety: isolate OK1-1

C;Date: 19-Mar-1997 #sequence\_revision 19-Mar-1997 #text\_change 05-Nov-1999

C;Accession: S49245

R;Chen, J.; MacFarlane, S.; Michael, T.; Wilson, A.

submitted to the EMBL Data Library, September 1994

A;Description: Deletion formation in the readthrough-protein gene from SBMWV RNA2.

A;Reference number: S49244

A;Accession: S49245

A;Molecule type: Genomic RNA

A;Residues: 1-174 <CHE>

A;Cross-references: EMBL:X81639; NID:g547536; PIDN:CAA57295.1; PID:g547538

A;Experimental source: isolate OK1-1; strain Oklahoma

C;Superfamily: soil-borne wheat mosaic virus hypothetical 19K protein

Query Match 29.7%; Score 56.5; DB 2; Length 174;  
Best Local Similarity 31.5%; Pred. No. 16;  
Matches 17; Conservative 7; Mismatches 11; Indels 19; Gaps 2;

QY 3 WLKARVEQ---ELQALEARGTDSNAELRAMEA-----KLKAEI 37

Db 84 WRHAQLRSLNAELDALEAREESRAQIKALSAGDHCPAVLAYVPKLTKLKAEV 137

## RESULT 14

E87478

GGDEF family protein [imported] - Caulobacter crescentus

C;Species: Caulobacter crescentus

C;Date: 20-Apr-2001 #sequence\_revision 20-Apr-2001 #text\_change 20-Apr-2001

C;Accession: E87478

R;Niernan, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J. B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolon n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M. Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001

A;Title: Complete Genome Sequence of Caulobacter crescentus.

A;Reference number: A87249; MUID:21173698; PMID:11259647

A;Accession: E87478

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-353 <STO>

A;Cross-references: GB:AE005673; NID:gi3423291; PIDN:AAK23825.1; GSPDB:GN00148

C;Genetics:

A;Gene: CC1850

Query Match 29.5%; Score 56; DB 2; Length 353;

Best Local Similarity 46.4%; Pred. No. 38;

Matches 13; Conservative 5; Mismatches 10; Indels 0; Gaps 0;

QY 7 RVEQELQALEARGTDSNAELRAMEAKLK 34

Db 147 RVHKNQSLAEARLADSTAEVRLREHLE 174

## RESULT 15

G71490

hypothetical protein CT622 - Chlamydia trachomatis (serotype D, strain UW3/Cx)

N;Alternate names: chlpn 76kda homolog CT622

C;Species: Chlamydia trachomatis

C;Date: 13-Sep-1998 #sequence\_revision 13-Sep-1998 #text\_change 08-Oct-1999

C;Accession: G71490

R;Stephens, R.S.; Kalman, S.; Lammel, C.J.; Fan, J.; Marathe, R.; Aravind, L.; Mitchell,

Science 282, 754-759, 1998

A;Title: Genome sequence of an obligate intracellular pathogen of humans: Chlamydia tract

A;Reference number: A71570; MUID:99000809; PMID:9784136

A;Accession: G71490

A;Molecule type: DNA

A;Residues: 1-647 <ARN>

A;Cross-references: GB:AE001333; GB:AE001273; NID:g3329068; PIDN:AAC68226.1; PID:g3329068

A;Experimental source: serotype D, strain UW-3/Cx

C;Comment: This sequence was originally identified as homologous to part of a sequence (s

PIR:E72042).

C;Genetics:

A;Gene: CT622

Query Match 29.5%; Score 56; DB 2; Length 647;

Best Local Similarity 33.3%; Pred. No. 71;

Matches 16; Conservative 9; Mismatches 13; Indels 10; Gaps 1;

QY 1 MDWLKARVEQ-----ELQALEARGTDSNAELRAMEAKLKAEIQ 38

Db 364 MQGFRSMIEQFNVPNPATAKELQAMEAQLTAMSDQLVGADGELPAEIQ 411

Search completed: January 30, 2004, 13:22:45

Job time : 22 secs

■ ■ ■



PT expressed in vivo in plants to serve as custom-tailored storage  
 PT proteins with specified aminoacid content

PS Disclosure; Page 117; 176pp; English.

XX The sequence represents a synthetic heptad polypeptide which can be  
 CC expressed in vivo in plants to serve as a synthetic seed storage  
 CC protein which can be custom-tailored for specific end-user requirements.  
 CC The DNA encoding the heptad may be used to transform plants to increase  
 CC the content of partic. amino acids such as lysine or methionine in seeds  
 CC or leaves. See also AAR31979-86 and AAR32659-705.  
 CC (Updated on 25-MAR-2003 to correct PN field.)

XX Sequence 49 AA;

Query Match 31.3%; Score 59.5; DB 14; Length 49;

Best Local Similarity 44.4%; Pred. No. 1.5;

Matches 16; Conservative 9; Mismatches 10; Indels 1; Gaps 1;

OY 4 LKARVQEQLQALRAGTDSNAELRAMEAKLKAEIQK 39  
 ||| :||:| :| :||| :||| :|  
 Db 12 LKA-MEEKLKAMEEKLKAMEEKLKAMEEK 46

#### RESULT 4

AAR32692  
 ID AAR32692 standard; Protein; 49 AA.

XX AAR32692;

XX 25-MAR-2003 (updated)

DT 22-JUN-1993 (first entry)

DE SSP polypeptide produced from clone 82-4.

XX Heptad; plants; custom tailored storage proteins; in vivo;  
 KW expression.

XX Synthetic.

XX WO9303160-A1.

PD 18-FEB-1993.

XX 07-AUG-1992; 92WO-US06412.

XX 09-AUG-1991; 91US-0743006.

XX (DUPO ) DU PONT DE NEMOURS & CO E I.

XX Falco SC, Keeler SJ, Rice JA;

XX WPI; 1993-076517/09.

DR N-PSDB; AAQ37273.

XX Synthetic polypeptide(s) contg. specified heptad units -  
 PT expressed in vivo in plants to serve as custom-tailored storage  
 PT proteins with specified aminoacid content

PS Disclosure; Page 123; 176pp; English.

XX The sequence represents a synthetic heptad polypeptide which can be  
 CC expressed in vivo in plants to serve as a synthetic seed storage  
 CC protein which can be custom-tailored for specific end-user requirements.  
 CC The DNA encoding the heptad may be used to transform plants to increase  
 CC the content of partic. amino acids such as lysine or methionine in seeds  
 CC or leaves. See also AAR31979-86 and AAR32659-705.  
 CC (Updated on 25-MAR-2003 to correct PN field.)

XX Sequence 49 AA;

Query Match 31.3%; Score 59.5; DB 14; Length 49;

Best Local Similarity 44.4%; Pred. No. 1.5;

Matches 16; Conservative 9; Mismatches 10; Indels 1; Gaps 1;

OY 4 LKARVQEQLQALRAGTDSNAELRAMEAKLKAEIQK 39  
 ||| :||:| :| :||| :||| :|  
 Db 5 LKA-MEEKLKAMEEKLKAMEEKLKAMEEK 39

#### RESULT 5

AAR78249

ID AAR78249 standard; Protein; 49 AA.

XX AAR78249;

XX 15-JUL-1996 (first entry)

XX SSP 7.7.7.7.7.5.

XX Lysine; synthetic storage protein; SSP; vector; PSK6;  
 KW dihydrodipicolinic acid synthase; corn; maize; Zea mays;  
 KW soybean; Glycine max; transgenic plant; essential amino acid.  
 XX Synthetic.

XX Location/Qualifiers

FT Key 1..7  
 FT Peptide /label= SSP7  
 FT Peptide 8..14  
 FT Peptide /label= SSP7  
 FT Peptide 15..21  
 FT Peptide /label= SSP7  
 FT Peptide 22..28  
 FT Peptide /label= SSP7  
 FT Peptide 28..35  
 FT Peptide /label= SSP7  
 FT Peptide 35..42  
 FT Peptide /label= SSP7  
 FT Peptide 43..49  
 FT Peptide /label= SSP5

XX WO9515392-A1.

XX 08-JUN-1995.

XX 21-NOV-1994; 94WO-US13190.

XX 17-JUN-1994; 94US-0261661.

PR 30-NOV-1993; 93US-0160117.

XX (DUPO ) DU PONT DE NEMOURS & CO E I.

XX Falco SC, Keeler SJ, Rice JA;

XX WPI; 1995-215272/28.

DR N-PSDB; AAQ94992.

XX New chimeric gene providing increased lysine content in plant seeds  
 PT - contains di:hydro:di:picolinic acid synthase gene coupled to  
 PT chloroplast transport sequence and seed specific promoter, also new  
 PT plants of improved nutritional value.

XX Example 8; Page 135; 180pp; English.

XX Lysine-rich synthetic storage protein SSP 7.7.7.7.7.5 (AAR78249)  
 CC comprises 6 heptad SSP7 repeats (see AAR78237) and a C-terminal SSP5  
 CC heptad (see AAR78236) derived from vector PSK6; an N-terminal SSP5  
 CC was lost owing to homologous recombination between base gene  
 CC repeats. The SSP was obt'd. by insertion of SSP7-encoding  
 CC oligonucleotides (AAQ94976-77) into PSK6 and use of the resulting vector  
 CC to transform Escherichia coli JM103, yielding clone 82-4 (AAQ94992).  
 CC The SSP forms a coiled-coil structure. It can be expressed in the  
 CC seeds of transformed plants, e.g. soybean and corn, to increase  
 CC lysine content.

XX





DD

DD



```

XX 29-MAR-1999 (first entry)
XX Synthetic lysine-rich storage protein 5.7.7.7.7.5.
DE Lysine; transgenic plant; seed storage protein.
XX Synthetic.
XX Key Location/Qualifiers
FH Peptide 1..7
FT Peptide /label= SSP5
FT Peptide 8..14
FT Peptide /label= SSP7
FT Peptide 15..21
FT Peptide /label= SSP7
FT Peptide 22..28
FT Peptide /label= SSP7
FT Peptide 28..35
FT Peptide /label= SSP7
FT Peptide 36..42
FT Peptide /label= SSP7
FT Peptide 43..49
FT Peptide /label= SSP5
XX WO9842831-A2.
PN 01-OCT-1998.
XX 27-MAR-1998; 98WO-US06051.
XX 27-MAR-1997; 97US-0824627.
XX (DUPO ) DU PONT DE NEMOURS & CO E I.
XX Epelbaum SU, Falco SC, McDevitt RE;
XX WPI; 1999-045139/04.
XX N-PSDB; AAV99514.
XX Nucleic acids and chimeric genes for increasing seed lysine content
PT - comprise sequence encoding all or part of lysine ketoglutarate
PT reductase, useful to improve nutritional quality of seeds from
PT transformed plants
XX Example 21; Page 135; 231pp; English.
XX This is the amino acid sequence of a lysine-rich synthetic seed
CC storage protein (SSP) composed of SSP5 and SSP7 heptad repeats
CC (sequence by heptad repeat is 5.7.7.7.7.5). A synthetic gene
CC (see AAV99514) encoding the SSP was constructed by insertion of
CC SSP7-encoding oligonucleotides (see AAV99509) into a unique EarI
CC site in the SSP5-encoding 'base gene' (see AAV99505) of vector pSK5.
CC Chimeric genes for lysine-rich synthetic seed storage proteins
CC suitable for expression in the seeds of plants have been
CC constructed. The invention provides methods for improving the
CC nutritional quality of seeds from transgenic plants by increasing
CC lysine content.
XX Sequence 49 AA;
SQ Query Match 31.3%; Score 59.5; DB 20; Length 49;
Best Local Similarity 44.4%; Pred. No. 1.5;
Matches 15; Conservative 9; Mismatches 10; Indels 1; Gaps 1;
Qy 4 LKARVEQLQALFARGTDSNAELRAMEAKLKAEIQK 39
Db 12 LKA-MEEKLKAMEEKLKAMEEKLKAMEEKLKAMEEK 46
RESULT 13
AAW87747
ID AAW87747 standard; Protein; 49 AA.

```

```

XX AAW87747;
XX 29-MAR-1999 (first entry)
XX Synthetic lysine-rich storage protein 7.7.7.7.7.5.
XX Lysine; transgenic plant; seed storage protein.
XX Synthetic.
XX Key Location/Qualifiers
FH Peptide 1..7
FT Peptide /label= SSP7
FT Peptide 8..14
FT Peptide /label= SSP7
FT Peptide 15..21
FT Peptide /label= SSP7
FT Peptide 22..28
FT Peptide /label= SSP7
FT Peptide 29..35
FT Peptide /label= SSP7
FT Peptide 36..42
FT Peptide /label= SSP7
FT Peptide 43..49
FT Peptide /label= SSP5
XX WO9842831-A2.
PN 01-OCT-1998.
XX 27-MAR-1998; 98WO-US06051.
XX 27-MAR-1997; 97US-0824627.
XX (DUPO ) DU PONT DE NEMOURS & CO E I.
XX Epelbaum SU, Falco SC, McDevitt RE;
XX WPI; 1999-045139/04.
XX N-PSDB; AAV99527.
XX Nucleic acids and chimeric genes for increasing seed lysine content
PT - comprise sequence encoding all or part of lysine ketoglutarate
PT reductase, useful to improve nutritional quality of seeds from
PT transformed plants
XX Example 21; Page 144; 231pp; English.
XX This is the amino acid sequence of a lysine-rich synthetic seed
CC storage protein (SSP) composed of SSP7 and SSP5 heptad repeats
CC (sequence by heptad repeat is 7.7.7.7.7.5). A synthetic gene
CC (see AAV99527) encoding the SSP was constructed by insertion of
CC SSP7-encoding oligonucleotides (see AAV99509) into a unique EarI
CC site in the SSP5-encoding 'base gene' (see AAV99505) of vector pSK6.
CC Chimeric genes for lysine-rich synthetic seed storage proteins
CC suitable for expression in the seeds of plants have been
CC constructed. The invention provides methods for improving the
CC nutritional quality of seeds from transgenic plants by increasing
CC lysine content.
XX Sequence 49 AA;
SQ Query Match 31.3%; Score 59.5; DB 20; Length 49;
Best Local Similarity 44.4%; Pred. No. 1.5;
Matches 16; Conservative 9; Mismatches 10; Indels 1; Gaps 1;
Qy 4 LKARVEQLQALFARGTDSNAELRAMEAKLKAEIQK 39
Db 5 LKA-MEEKLKAMEEKLKAMEEKLKAMEEKLKAMEEK 39
RESULT 14

```

